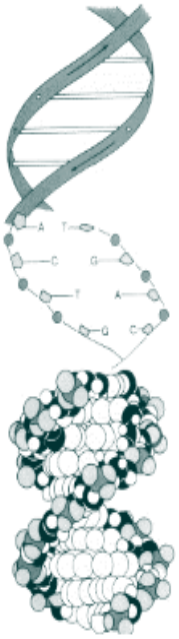


Instrumenting the Health Care Enterprise for Discovery in the Course of Clinical Care

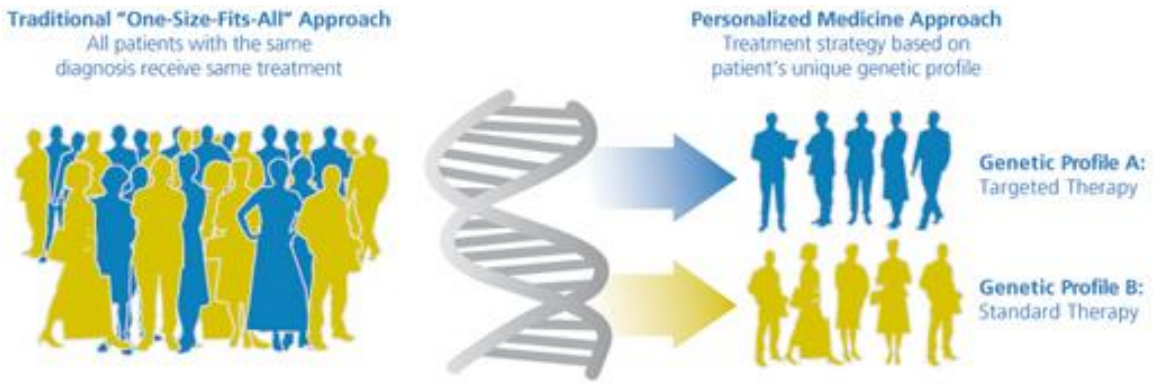
Shawn Murphy MD, Ph.D.

Professor of Neurology and Biomedical Informatics

Harvard Medical School

Chief Research Information Officer, Partners Healthcare

Personalized Medicine and Genomic technology are critical to managing populations

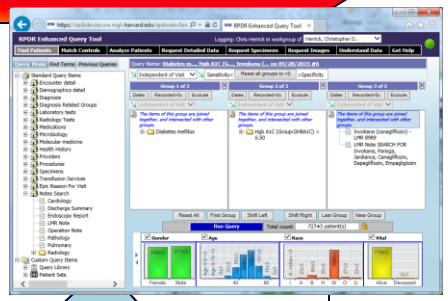


- Managing a population involves improving health outcomes of the group as a whole by identifying, monitoring and addressing health needs of individuals through:
  - Subpopulation stratification
  - Targeted, evidence-based treatment protocols
  - Predictive analytics

Source: Personalized Medicine Coalition and innovation.org ; Oliver Wyman

The RPDR Warehouse at Partners Healthcare

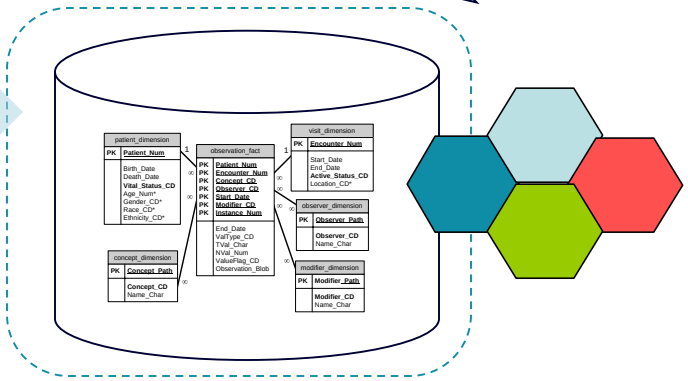
Partners Enterprise



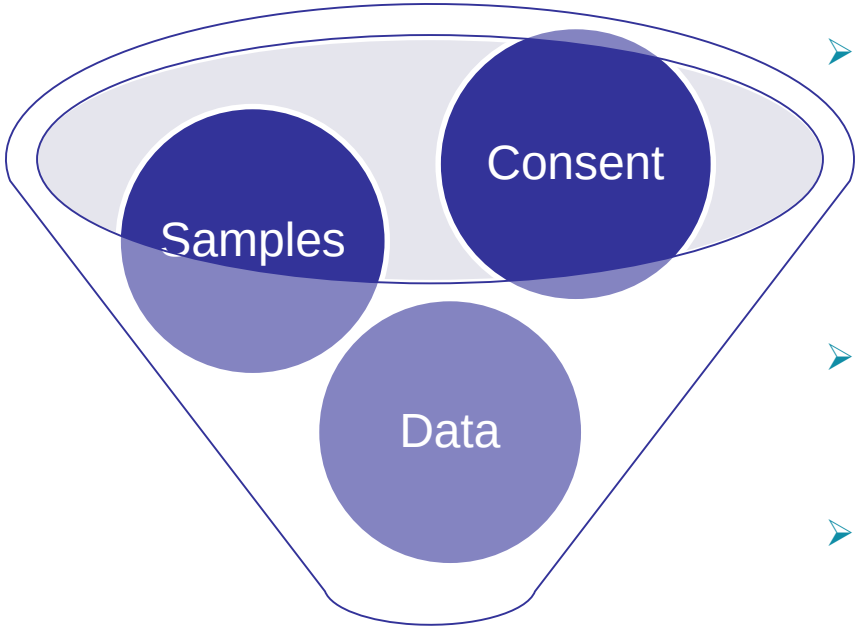
Research Patient Data Registry

Multiple Systems at Partners:

- Billing Data
- Epic Data
- Research Data (consent to contact)
- Specimen Data
- Laboratory Data
- …



The Partners Biobank



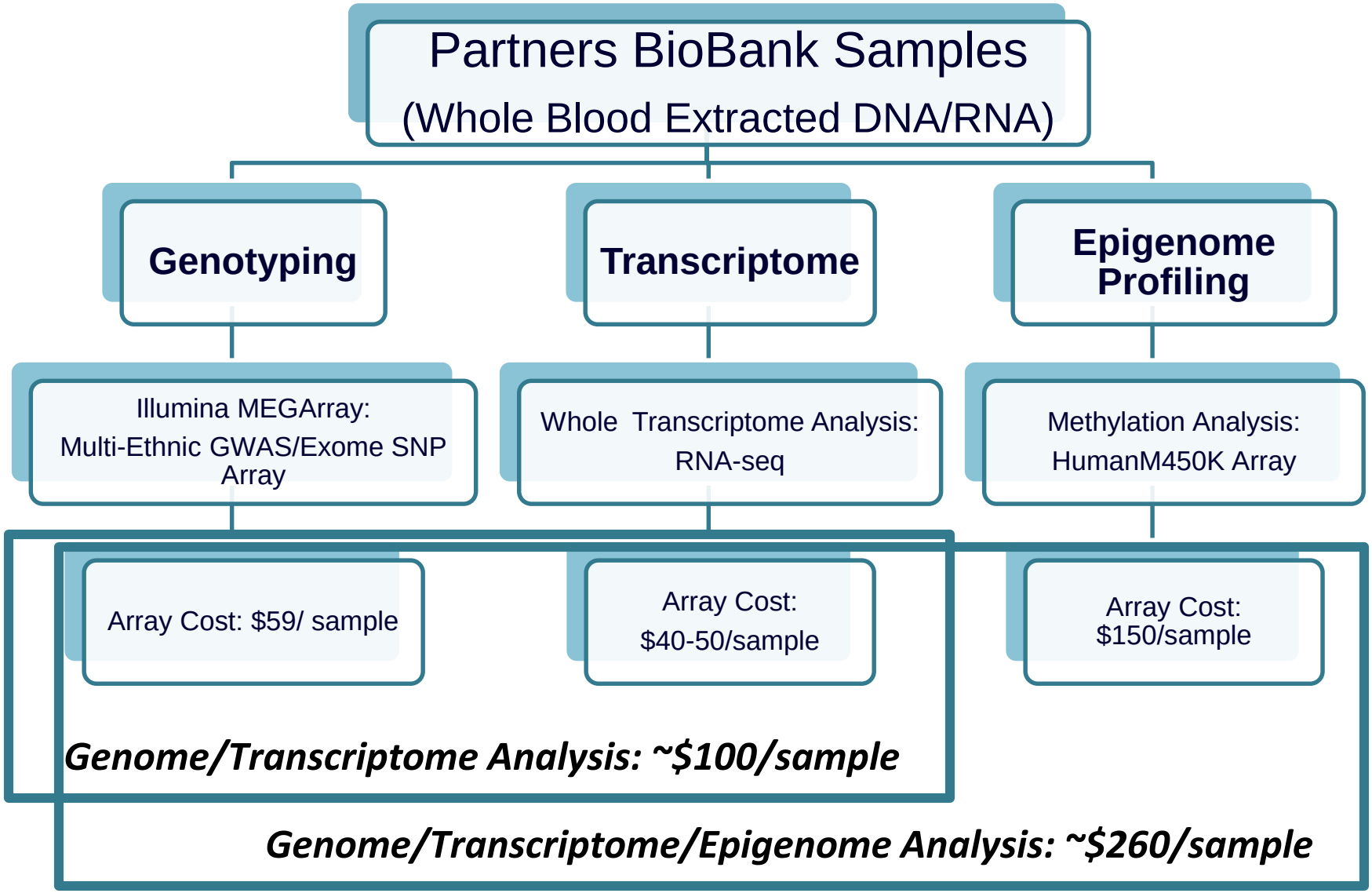
Research Discoveries

- The Partners Biobank provides samples (plasma, serum, and DNA) collected from consented patients.
- 84,000 patients have consented to date
- Samples are available for distribution to Partners investigators* to help identify novel Personalized Medicine opportunities that reduce cost and provide better care

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

Improved Clinical Care for All Patients

Biobank Integrative Genomics Strategy

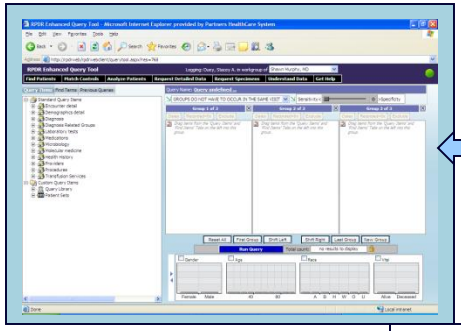


Research Patient Data Registry (RPDR) to find patient cohorts and distribute data

1) Queries for aggregate patient numbers

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

Query construction in web tool



De-identified Data Warehouse

Z731984X
Z74902XX
...
...

Encrypted identifiers

2) Returns detailed patient data

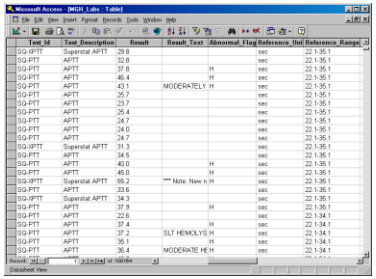
- Start with list of specific patients, usually from (1)
- Authorized use by IRB Protocol
- Returns contact and PCP information, demographics, providers, visits, diagnoses, medications, procedures, laboratories, microbiology, reports (discharge, LMR, operative, radiology, pathology, cardiology, pulmonary,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

0000004
2185793
...
...

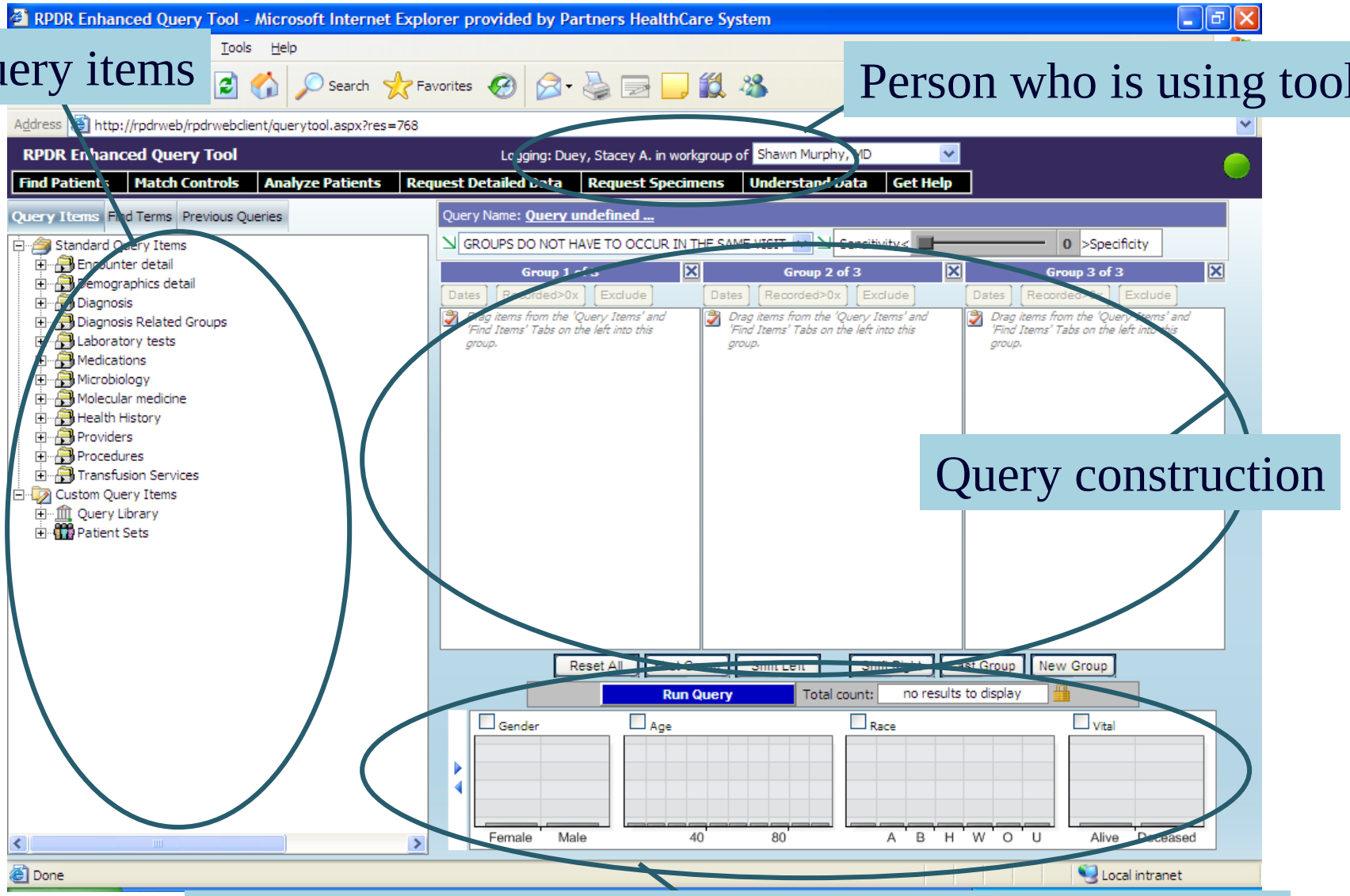
OR

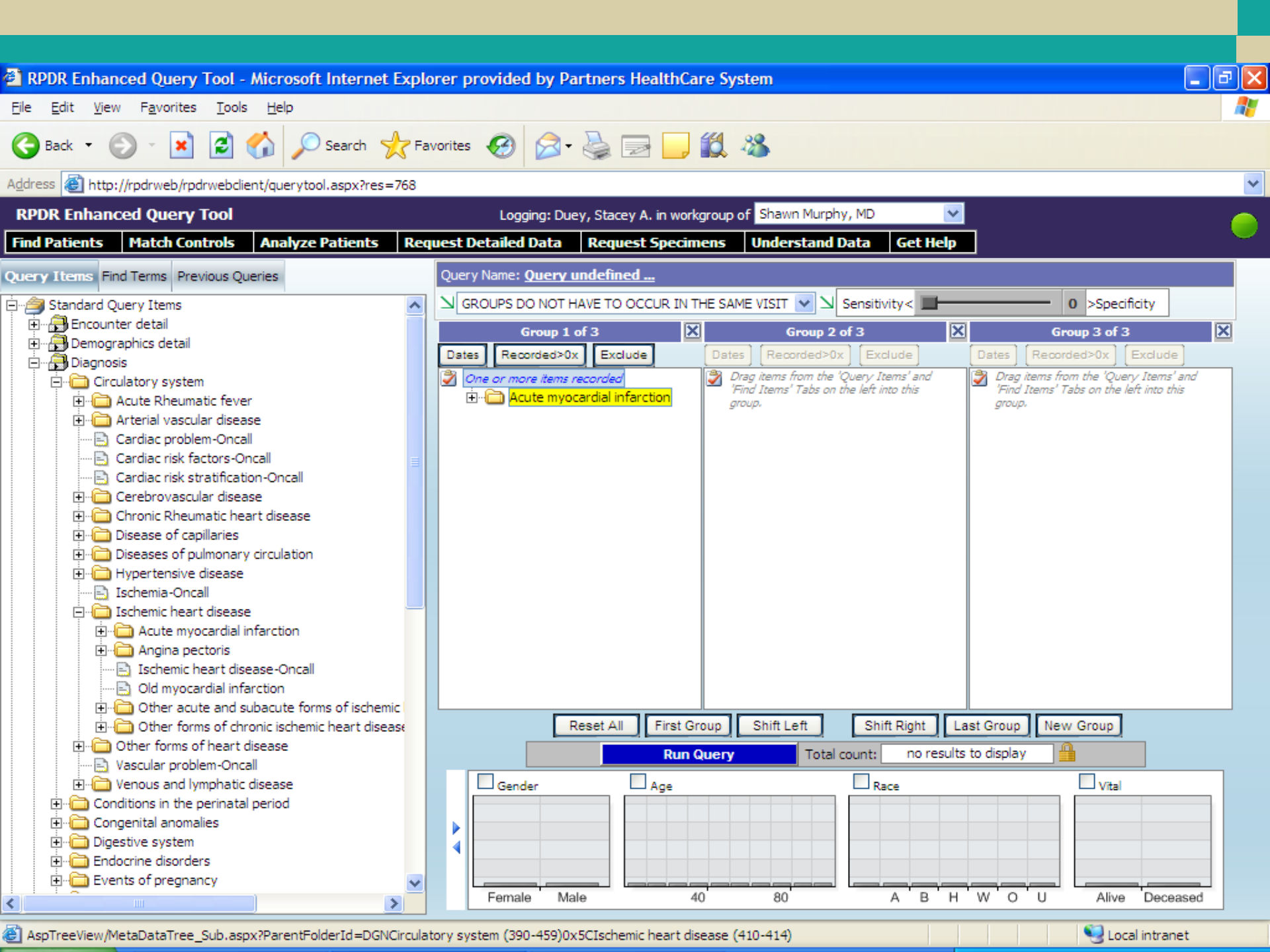
0000004
2185793
...
...

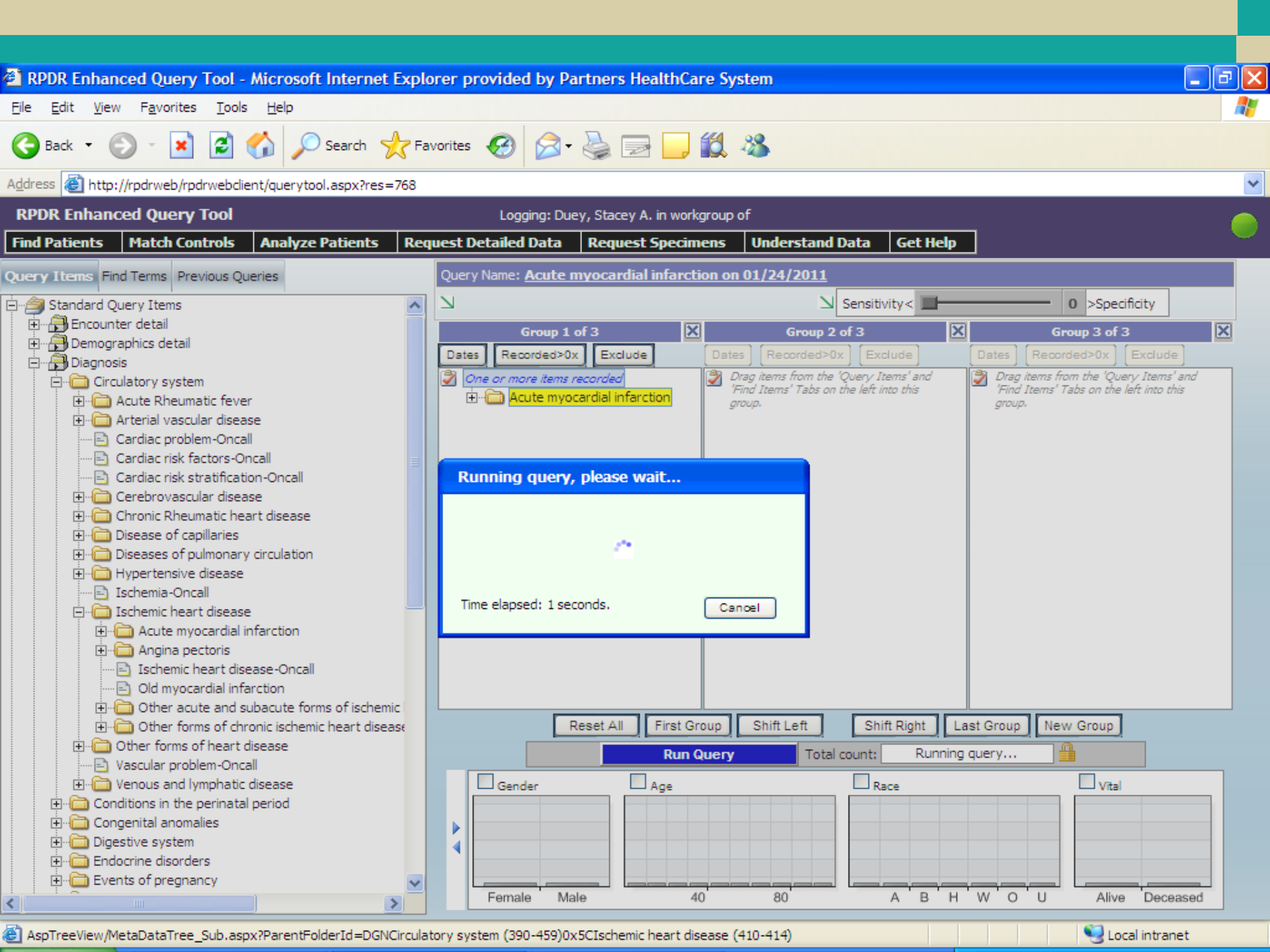
Real identifiers

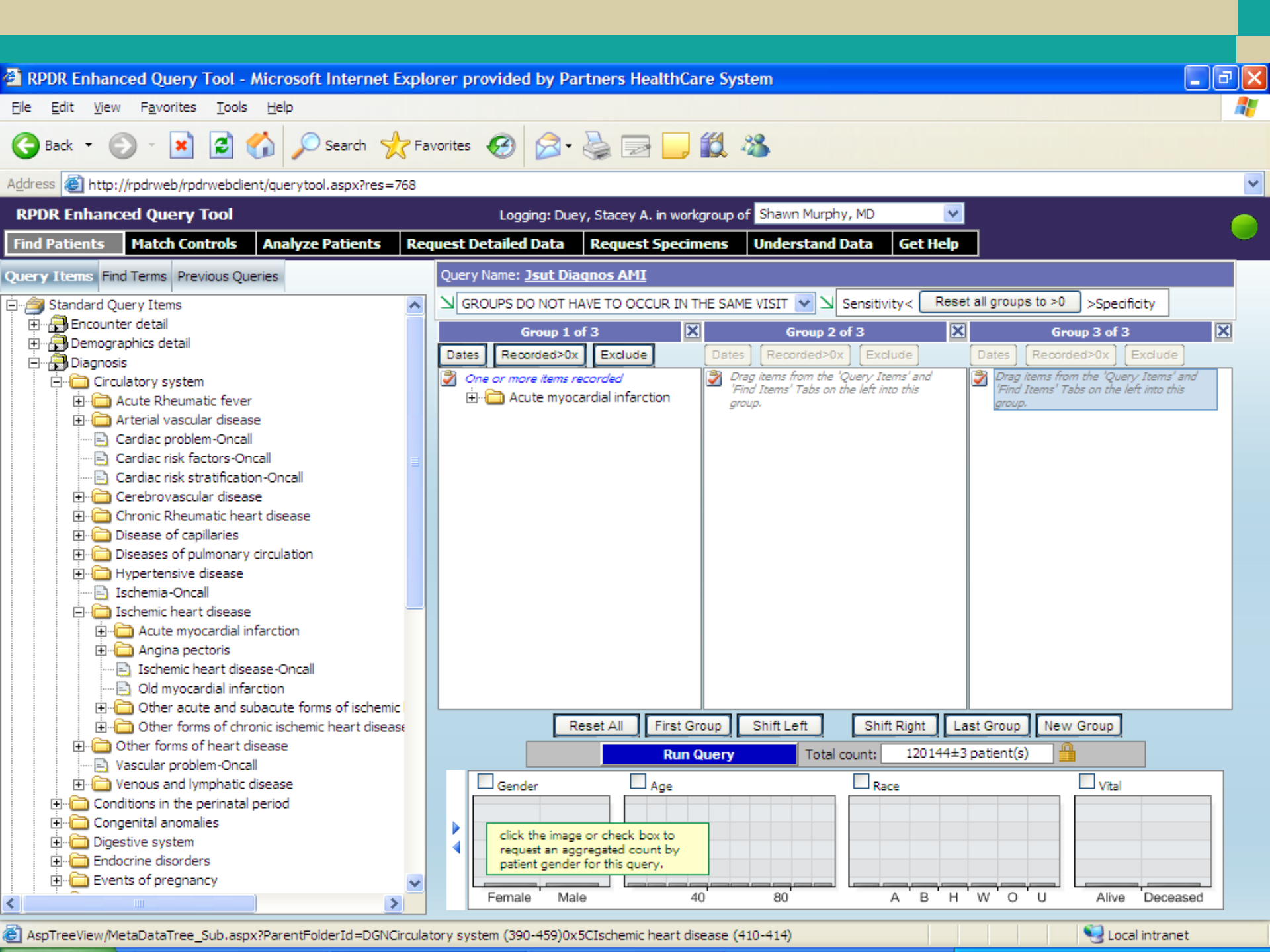


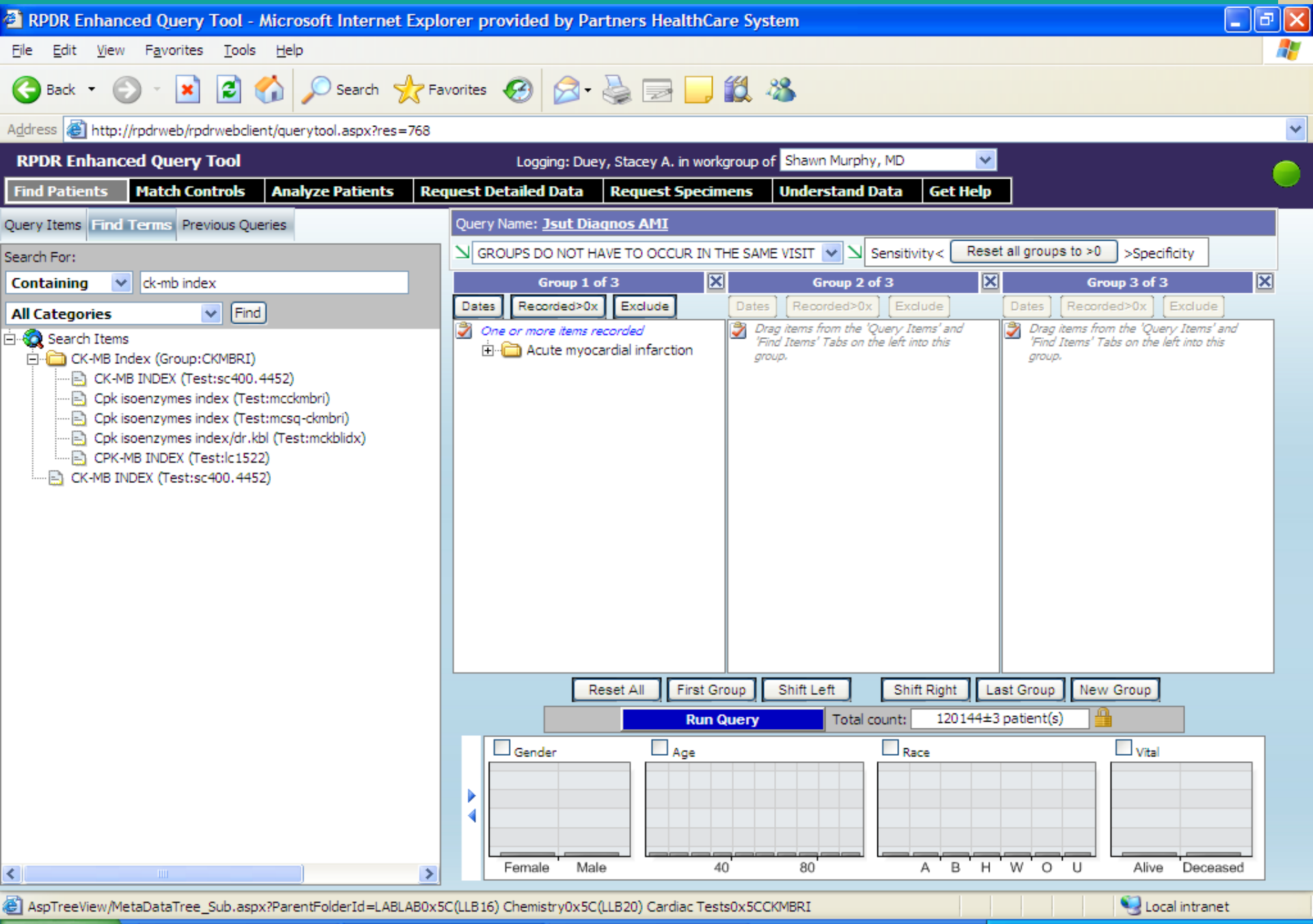
FINDING PATIENTS

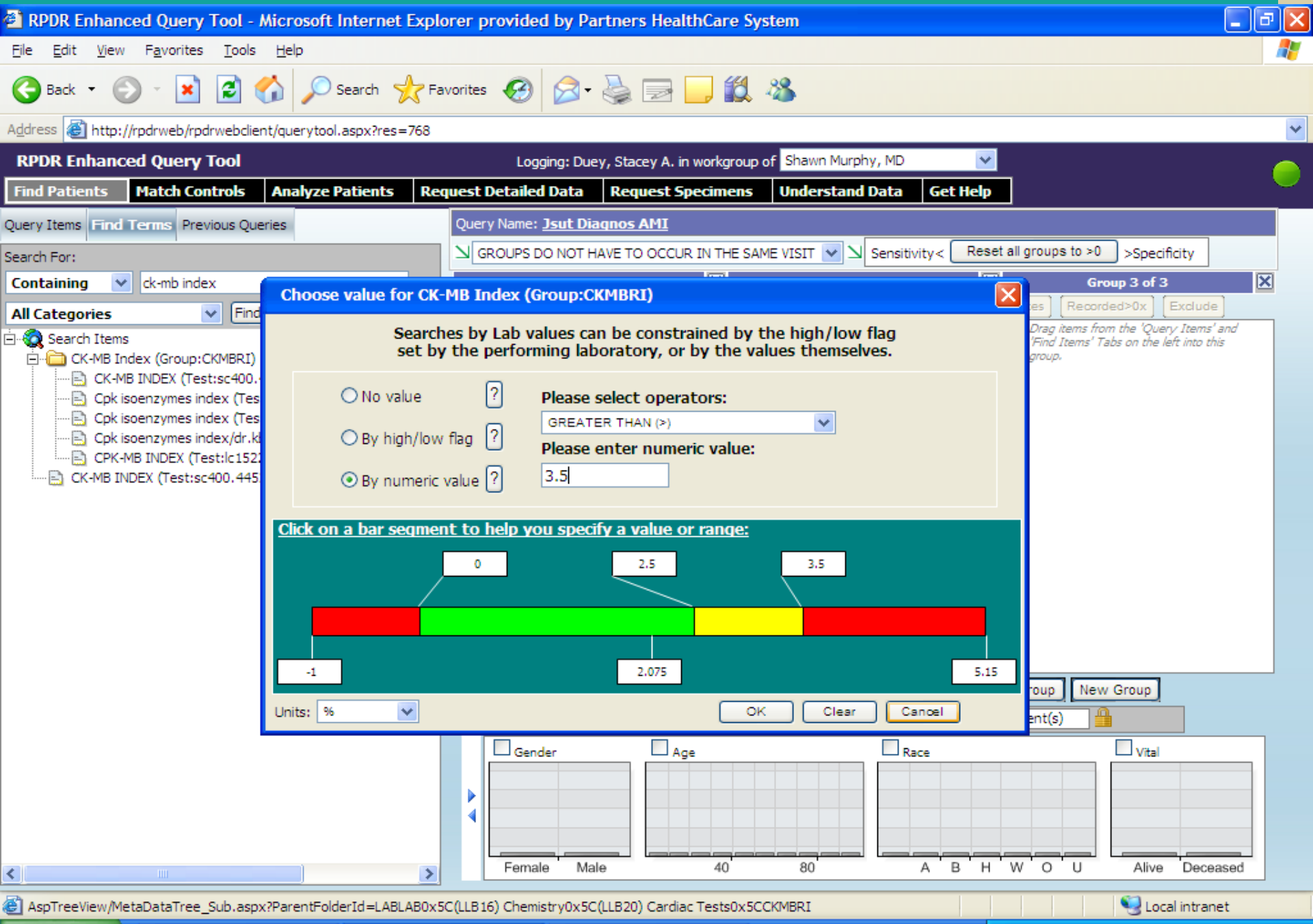


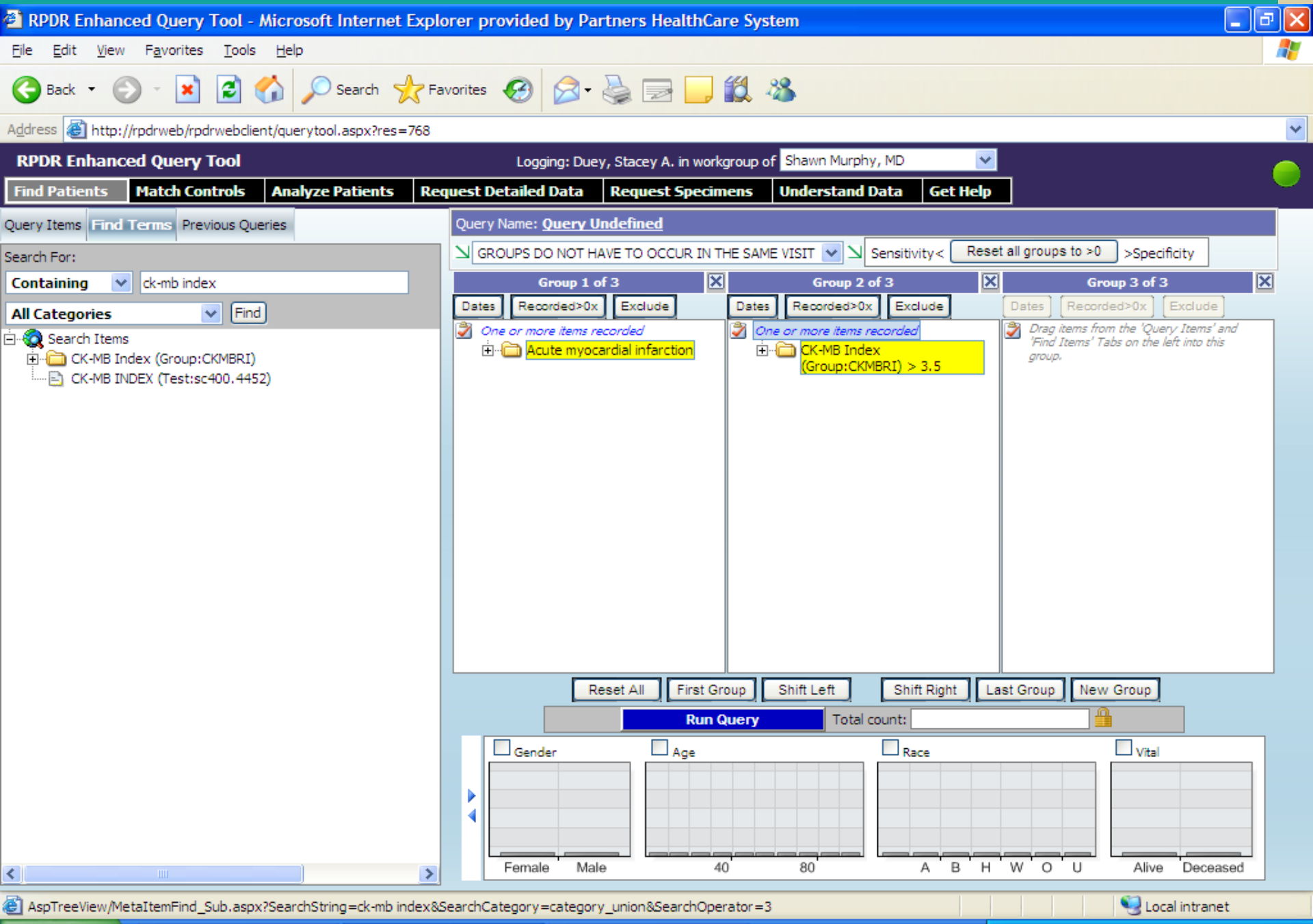

RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System
File Edit View Favorites Tools Help
Back Search Favorites
Address http://rpdrweb/rpdrwebclient/querytool.aspx?res=768
RPDR Enhanced Query Tool
Logging: Duey, Stacey A. in workgroup of Shawn Murphy, MD
Find Patients
Match Controls
Analyze Patients
Request Detailed Data
Request Specimens
Understand Data
Get Help
Query Items
Find Terms
Previous Queries
Standard Query Items
Encounter detail
Demographics detail
Diagnosis
Circulatory system
Acute Rheumatic fever
Arterial vascular disease
Cardiac problem-Oncall
Cardiac risk factors-Oncall
Cardiac risk stratification-Oncall
Cerebrovascular disease
Chronic Rheumatic heart disease
Disease of capillaries
Diseases of pulmonary circulation
Hypertensive disease
Ischemia-Oncall
Ischemic heart disease
Acute myocardial infarction
Angina pectoris
Ischemic heart disease-Oncall
Old myocardial infarction
Other acute and subacute forms of ischemic
Other forms of chronic ischemic heart disease
Other forms of heart disease
Vascular problem-Oncall
Venous and lymphatic disease
Conditions in the perinatal period
Congenital anomalies
Digestive system
Endocrine disorders
Events of pregnancy
Query Name: Query undefined ...
GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT
Sensitivity< 0 >Specificity
Group 1 of 3
Group 2 of 3
Group 3 of 3
Dates
Recorded>0x
Exclude
One or more items recorded
Acute myocardial infarction
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:
no results to display
Gender
Age
Race
Vital
Female Male
40 80
A B H W O U
Alive Deceased
AspTreeView/MetaDataTree_Sub.aspx?ParentFolderId=DGNCirculatory system (390-459)0x5CIschemic heart disease (410-414)
Local intranet

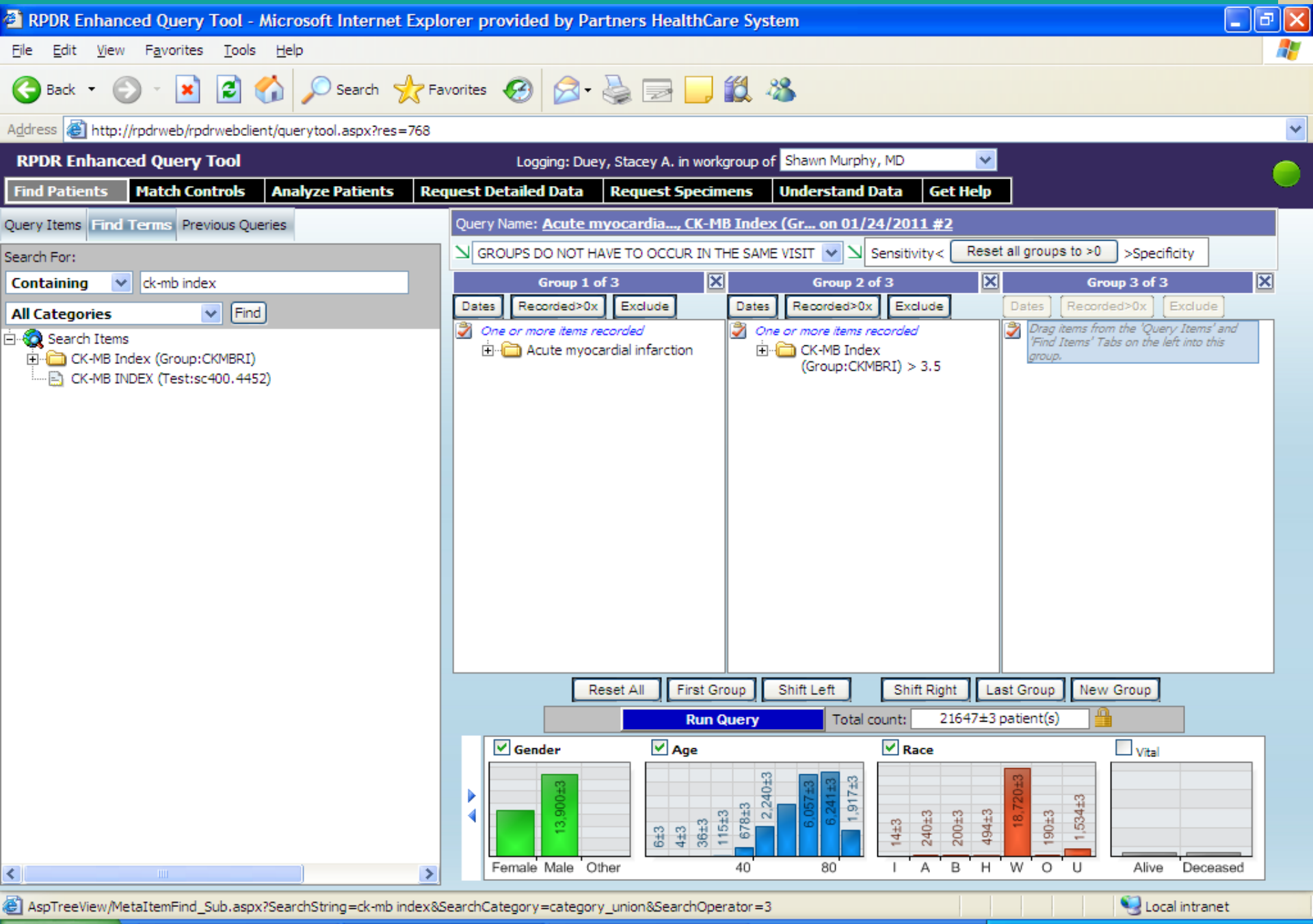

RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System
File Edit View Favorites Tools Help
Back Search Favorites
Address http://rpdrweb/rpdrwebclient/querytool.aspx?res=768
RPDR Enhanced Query Tool
Logging: Duey, Stacey A. in workgroup of
Find Patients
Match Controls
Analyze Patients
Request Detailed Data
Request Specimens
Understand Data
Get Help
Query Items
Find Terms
Previous Queries
Standard Query Items
Encounter detail
Demographics detail
Diagnosis
Circulatory system
Acute Rheumatic fever
Arterial vascular disease
Cardiac problem-Oncall
Cardiac risk factors-Oncall
Cardiac risk stratification-Oncall
Cerebrovascular disease
Chronic Rheumatic heart disease
Disease of capillaries
Diseases of pulmonary circulation
Hypertensive disease
Ischemia-Oncall
Ischemic heart disease
Acute myocardial infarction
Angina pectoris
Ischemic heart disease-Oncall
Old myocardial infarction
Other acute and subacute forms of ischemic
Other forms of chronic ischemic heart disease
Other forms of heart disease
Vascular problem-Oncall
Venous and lymphatic disease
Conditions in the perinatal period
Congenital anomalies
Digestive system
Endocrine disorders
Events of pregnancy
Query Name: Acute myocardial infarction on 01/24/2011
Sensitivity< 0 >Specificity
Group 1 of 3
Group 2 of 3
Group 3 of 3
Dates
Recorded>0x
Exclude
One or more items recorded
Acute myocardial infarction
Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group.
Running query, please wait...
Time elapsed: 1 seconds.
Cancel
Reset All
First Group
Shift Left
Shift Right
Last Group
New Group
Run Query
Total count:
Running query...
Gender
Age
Race
Vital
Female Male
40 80
A B H W O U
Alive Deceased
AspTreeView/MetaDataTree_Sub.aspx?ParentFolderId=DGNCirculatory system (390-459)0x5CIschemic heart disease (410-414)
Local intranet


RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System
File Edit View Favorites Tools Help
Back Search Favorites
Address http://rpdrweb/rpdrwebclient/querytool.aspx?res=768
RPDR Enhanced Query Tool
Logging: Duey, Stacey A. in workgroup of Shawn Murphy, MD
Find Patients
Match Controls
Analyze Patients
Request Detailed Data
Request Specimens
Understand Data
Get Help
Query Items
Find Terms
Previous Queries
Standard Query Items
Encounter detail
Demographics detail
Diagnosis
Circulatory system
Acute Rheumatic fever
Arterial vascular disease
Cardiac problem-Oncall
Cardiac risk factors-Oncall
Cardiac risk stratification-Oncall
Cerebrovascular disease
Chronic Rheumatic heart disease
Disease of capillaries
Diseases of pulmonary circulation
Hypertensive disease
Ischemia-Oncall
Ischemic heart disease
Acute myocardial infarction
Angina pectoris
Ischemic heart disease-Oncall
Old myocardial infarction
Other acute and subacute forms of ischemic
Other forms of chronic ischemic heart disease
Other forms of heart disease
Vascular problem-Oncall
Venous and lymphatic disease
Conditions in the perinatal period
Congenital anomalies
Digestive system
Endocrine disorders
Events of pregnancy
Query Name: Jsut Diagnos AMI
GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT
Sensitivity< Reset all groups to >0 >Specificity
Group 1 of 3
Group 2 of 3
Group 3 of 3
Dates
Recorded>0x
Exclude
One or more items recorded
Acute myocardial infarction
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
AspTreeView/MetaDataTree_Sub.aspx?ParentFolderId=DGNCirculatory system (390-459)0x5CIschemic heart disease (410-414)
Local intranet


RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System
File Edit View Favorites Tools Help
Back Search Favorites
Address http://rpdrweb/rpdrwebclient/querytool.aspx?res=768
RPDR Enhanced Query Tool
Logging: Duey, Stacey A. in workgroup of Shawn Murphy, MD
Find Patients
Match Controls
Analyze Patients
Request Detailed Data
Request Specimens
Understand Data
Get Help
Query Items
Find Terms
Previous Queries
Search For:
Containing
ck-mb index
All Categories
Find
Search Items
CK-MB Index (Group:CKMBRI)
CK-MB INDEX (Test:sc400.4452)
Cpk isoenzymes index (Test:mcckmbri)
Cpk isoenzymes index (Test:mcsq-ckmbri)
Cpk isoenzymes index/dr.kbl (Test:mckblidx)
CPK-MB INDEX (Test:lc1522)
CK-MB INDEX (Test:sc400.4452)
Query Name: Jsut Diagnos AMI
GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT
Sensitivity<
Reset all groups to >0
>Specificity
Group 1 of 3
Group 2 of 3
Group 3 of 3
Dates
Recorded>0x
Exclude
One or more items recorded
Acute myocardial infarction
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
Female
Male
40
80
A
B
H
W
O
U
Alive
Deceased
AspTreeView/MetaDataTree_Sub.aspx?ParentFolderId=LABLAB0x5C(LLB16) Chemistry0x5C(LLB20) Cardiac Tests0x5CCKMBRI
Local intranet


RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System
File Edit View Favorites Tools Help
Back Search Favorites
Address http://rpdrweb/rpdrwebclient/querytool.aspx?res=768
RPDR Enhanced Query Tool
Logging: Duey, Stacey A. in workgroup of Shawn Murphy, MD
Find Patients
Match Controls
Analyze Patients
Request Detailed Data
Request Specimens
Understand Data
Get Help
Query Items
Find Terms
Previous Queries
Search For:
Containing
ck-mb index
All Categories
Search Items
CK-MB Index (Group:CKMBRI)
CK-MB INDEX (Test:sc400.
Cpk isoenzymes index (Tes
Cpk isoenzymes index (Tes
Cpk isoenzymes index/dr.k
CPK-MB INDEX (Test:lc152
CK-MB INDEX (Test:sc400.445
Query Name: Jsut Diagnos AMI
GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT
Sensitivity<
Reset all groups to >0
>Specificity
Group 3 of 3
Recorded>0x
Exclude
Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group.
New Group
Choose value for CK-MB Index (Group:CKMBRI)
Searches by Lab values can be constrained by the high/low flag set by the performing laboratory, or by the values themselves.
No value
By high/low flag
By numeric value
Please select operators:
GREATER THAN (>)
Please enter numeric value:
3.5
Click on a bar segment to help you specify a value or range:
0
2.5
3.5
-1
2.075
5.15
Units: %
OK
Clear
Cancel
Gender
Female Male
Age
40 80
Race
A B H W O U
Vital
Alive Deceased
AspTreeView/MetaDataTree_Sub.aspx?ParentFolderId=LABLAB0x5C(LLB16) Chemistry0x5C(LLB20) Cardiac Tests0x5CCKMBRI
Local intranet


RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System
File Edit View Favorites Tools Help
Back Search Favorites
Address http://rpdrweb/rpdrwebclient/querytool.aspx?res=768
RPDR Enhanced Query Tool
Logging: Duey, Stacey A. in workgroup of Shawn Murphy, MD
Find Patients
Match Controls
Analyze Patients
Request Detailed Data
Request Specimens
Understand Data
Get Help
Query Items
Find Terms
Previous Queries
Search For:
Containing
ck-mb index
All Categories
Find
Search Items
CK-MB Index (Group:CKMBRI)
CK-MB INDEX (Test:sc400.4452)
Query Name: Query Undefined
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
One or more items recorded
Acute myocardial infarction
CK-MB Index (Group:CKMBRI) > 3.5
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:
Gender
Age
Race
Vital
Female
Male
40
80
A
B
H
W
O
U
Alive
Deceased
AspTreeView/MetaItemFind_Sub.aspx?SearchString=ck-mb index&SearchCategory=category_union&SearchOperator=3
Local intranet


RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System
File Edit View Favorites Tools Help
Back Search Favorites
Address http://rpdrweb/rpdrwebclient/querytool.aspx?res=768
RPDR Enhanced Query Tool
Logging: Duey, Stacey A. in workgroup of Shawn Murphy, MD
Find Patients
Match Controls
Analyze Patients
Request Detailed Data
Request Specimens
Understand Data
Get Help
Query Items
Find Terms
Previous Queries
Search For:
Containing
ck-mb index
All Categories
Find
Search Items
CK-MB Index (Group:CKMBRI)
CK-MB INDEX (Test:sc400.4452)
Query Name: Acute myocardia..., CK-MB Index (Gr... on 01/24/2011 #2
GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT
Sensitivity<
Reset all groups to >0
>Specificity
Group 1 of 3
Group 2 of 3
Group 3 of 3
Dates
Recorded>0x
Exclude
One or more items recorded
Acute myocardial infarction
CK-MB Index
(Group:CKMBRI) > 3.5
Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group.
Reset All
First Group
Shift Left
Shift Right
Last Group
New Group
Run Query
Total count:
21647±3 patient(s)
Gender
Female
Male
Other
13,900±3
Age
6±3
4±3
36±3
115±3
678±3
2,240±3
6,057±3
6,241±3
1,917±3
40
80
Race
14±3
240±3
200±3
494±3
18,720±3
190±3
1,534±3
I
A
B
H
W
O
U
Vital
Alive
Deceased
AspTreeView/MetaItemFind_Sub.aspx?SearchString=ck-mb index&SearchCategory=category_union&SearchOperator=3
Local intranet

RPDR Detailed Data Request Wizard -- Web Page Dialog

Welcome to the RPDR Data Request Wizards

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

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

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

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

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

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

I accept responsibility for the data returned by this query.

Accept **Cancel**

Partners Healthcare System HIPAA Compliance

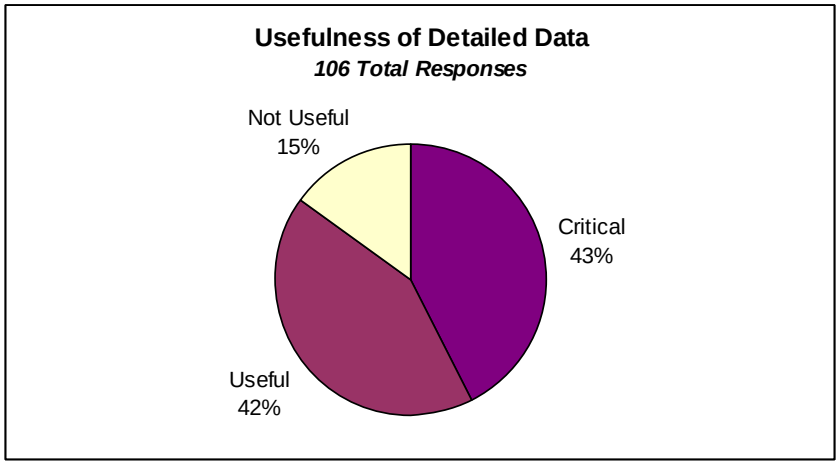
Additional HIPAA information for the research community is available from these links, sponsored by Partners and the Human Research Council (PHRC).

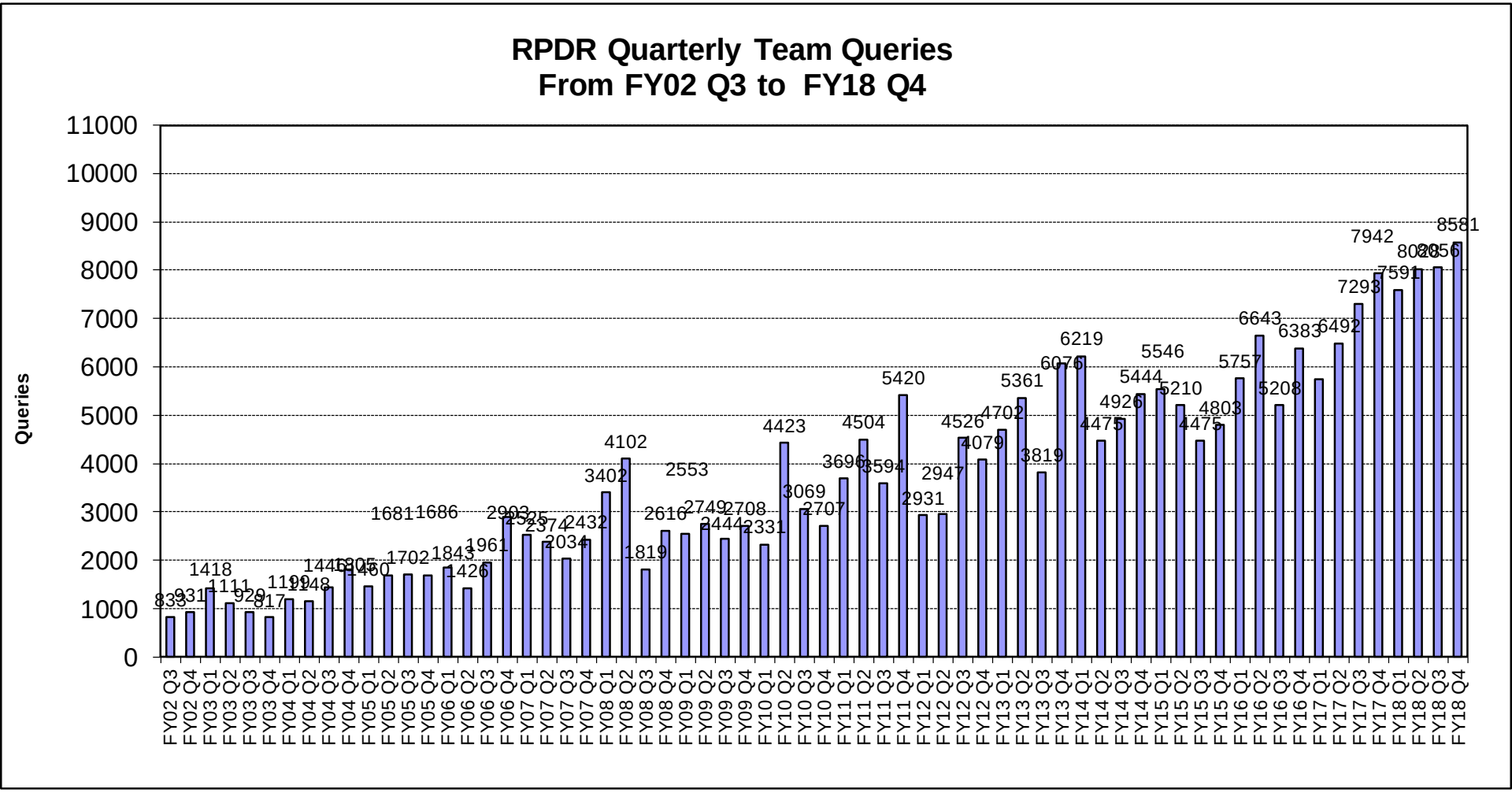
HIPAA and the Privacy Rule

HIPAA Central

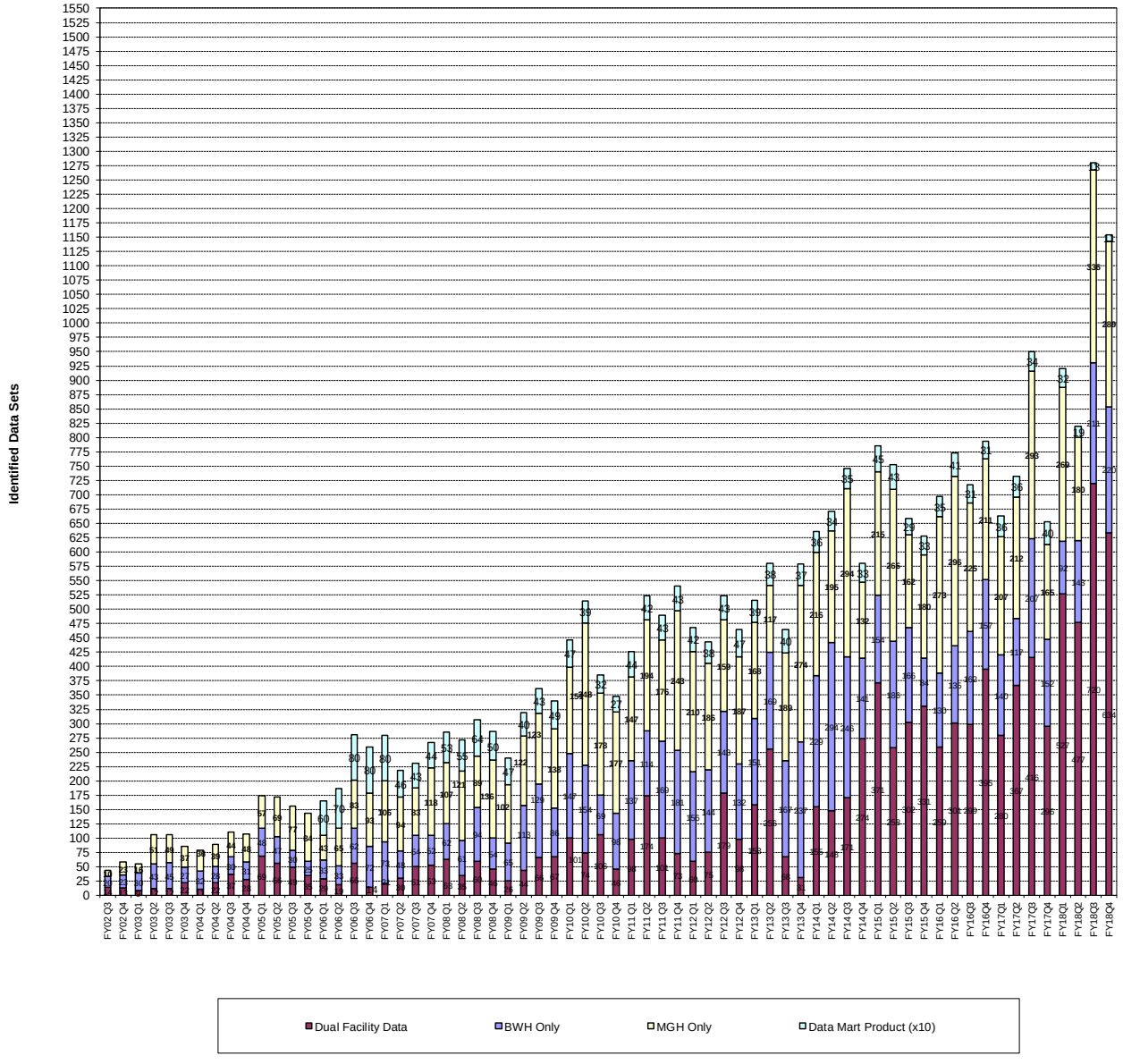
One year (2014) usage of RPDR

- 3100 registered users, 655 new in 2014
- 583 teams/year gathering data for research studies
- 2634 detailed patient data sets returned to these teams in 2014, containing data of 24.7 million patient records.
- From a survey of 153 teams
  - Importance of the data received from the RPDR was evaluated in relation to the study it was supporting.
  - $94-136 million/year total research support critically dependent on RPDR from patient data received throughout life of funding.
- ~300 data marts were created to support hospital operations, representing about 80 million patient records

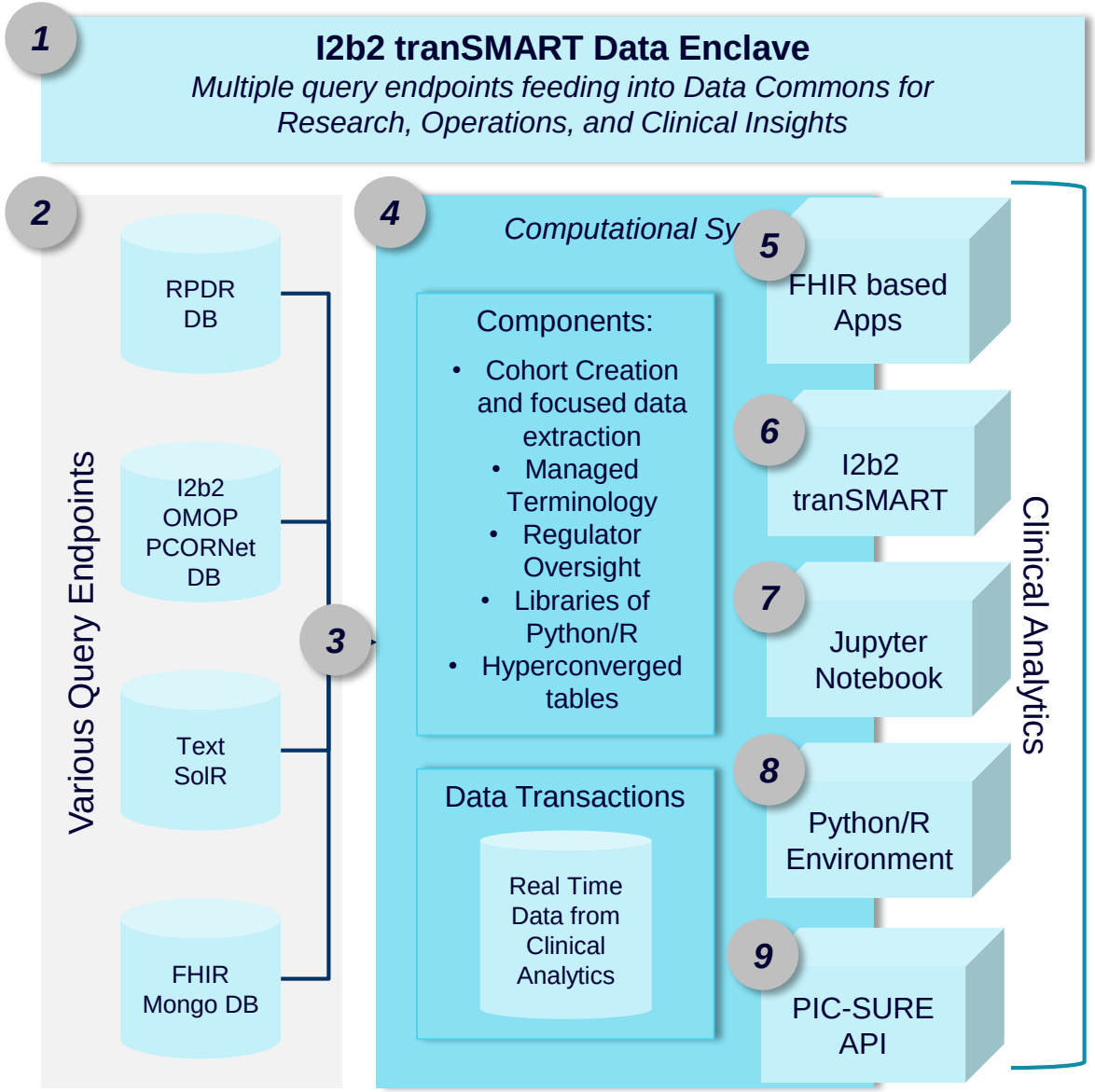



RPDR Quarterly Team Queries
From FY02 Q3 to FY18 Q4
Queries
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
FY02 Q3
FY02 Q4
FY03 Q1
FY03 Q2
FY03 Q3
FY03 Q4
FY04 Q1
FY04 Q2
FY04 Q3
FY04 Q4
FY05 Q1
FY05 Q2
FY05 Q3
FY05 Q4
FY06 Q1
FY06 Q2
FY06 Q3
FY06 Q4
FY07 Q1
FY07 Q2
FY07 Q3
FY07 Q4
FY08 Q1
FY08 Q2
FY08 Q3
FY08 Q4
FY09 Q1
FY09 Q2
FY09 Q3
FY09 Q4
FY10 Q1
FY10 Q2
FY10 Q3
FY10 Q4
FY11 Q1
FY11 Q2
FY11 Q3
FY11 Q4
FY12 Q1
FY12 Q2
FY12 Q3
FY12 Q4
FY13 Q1
FY13 Q2
FY13 Q3
FY13 Q4
FY14 Q1
FY14 Q2
FY14 Q3
FY14 Q4
FY15 Q1
FY15 Q2
FY15 Q3
FY15 Q4
FY16 Q1
FY16 Q2
FY16 Q3
FY16 Q4
FY17 Q1
FY17 Q2
FY17 Q3
FY17 Q4
FY18 Q1
FY18 Q2
FY18 Q3
FY18 Q4
835
931
1418
1111
929
817
1199
1148
1445
1805
1460
1702
1681
1686
1843
1426
1961
2903
2525
2374
2034
2432
3402
4102
1819
2616
2553
2749
2708
2444
2331
4423
3069
2707
3696
3594
4504
5420
2931
2947
4526
4079
4702
5361
3819
6076
6219
4475
4926
5444
5546
5210
4475
4803
5757
5208
6643
6383
6492
7293
7942
7591
8028
8056
8581

RPDR Detailed Data Set Production
FY 02 Q3 to FY 18 Q4

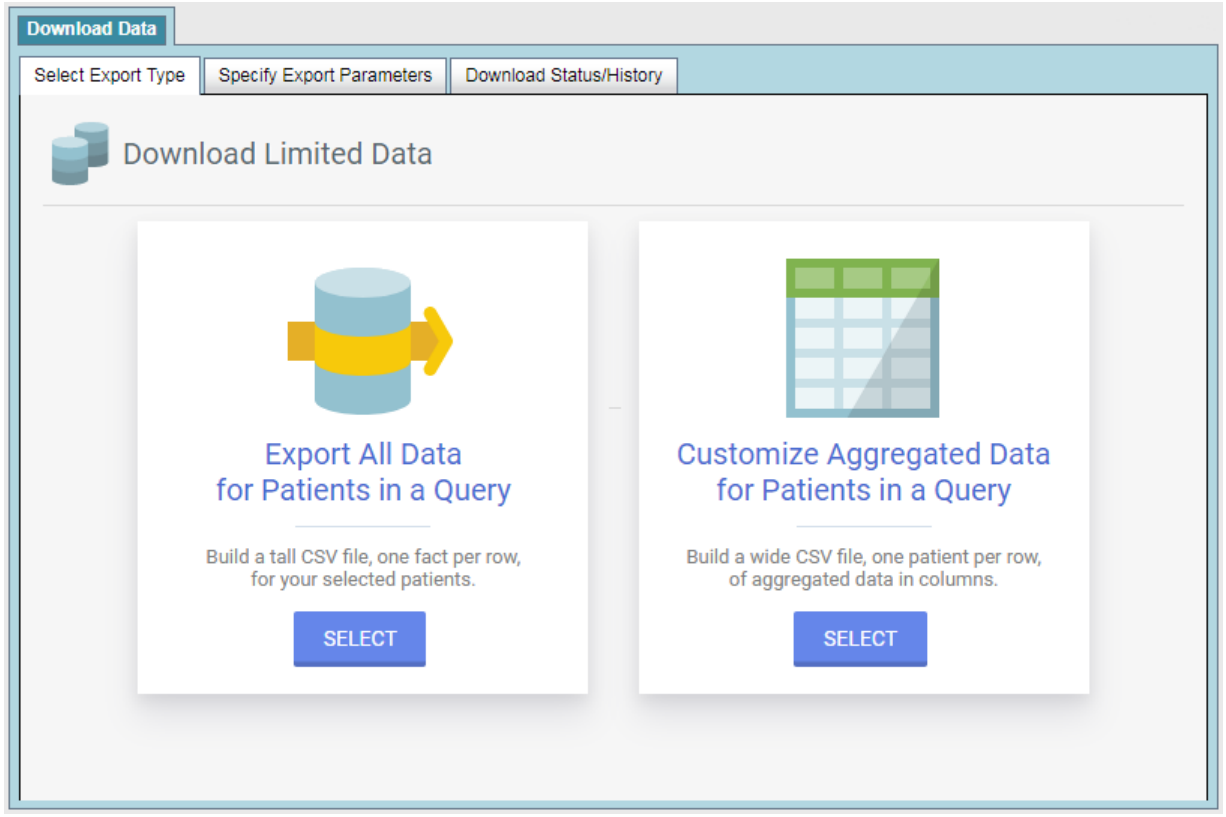


Secure Data Computation in Data Enclave

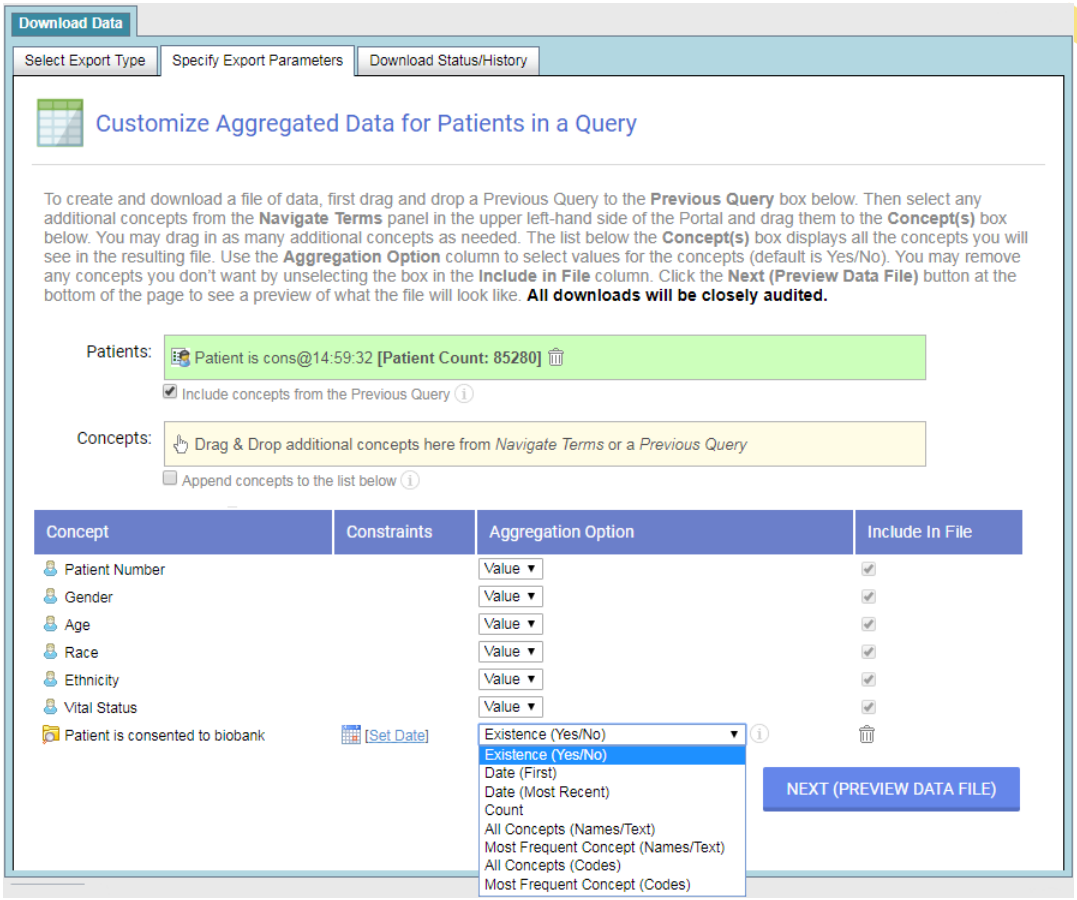


System Components	
1	Combining Big Data Sources
2	Query Endpoints for DATA
3	Web Services / Bulk Transfer
Technical Solution Development	
4	AI/ML Ready Bundle
Research and Clinical Application Projects	
5	FHIR based SMART Apps
6	I2b2 tranSMART with Fractalis plugin (next version of SmartR plugin)
7	Jupyter Notebook with AI Visualizations - Matching patients with similar disease profiles
8	Python/R Environment full interactive development

Hyperconverged Data Export Types



Sparse Matrix Tables for Computation



Patient Num	Gender	Age	Race	Ethnicity	Vital Status	Consented
1	M	10	W	Hispanic	Living	Yes
2	F	11	W	Hispanic	Living	Yes
3	M	12	B	Hispanic	Living	No
4	F	13	W	Hispanic	Living	No
5	M	14	B	Hispanic	Living	Yes
6	F	15	W	Hispanic	Living	Yes
7	M	16	B	Hispanic	Living	No

Key for Success – Two types of tables

[1] Sparse Matrix

Semantic (Ontology) value

Patients from Cohort selection

View table of patients

You are viewing rows 1-50 out of 251

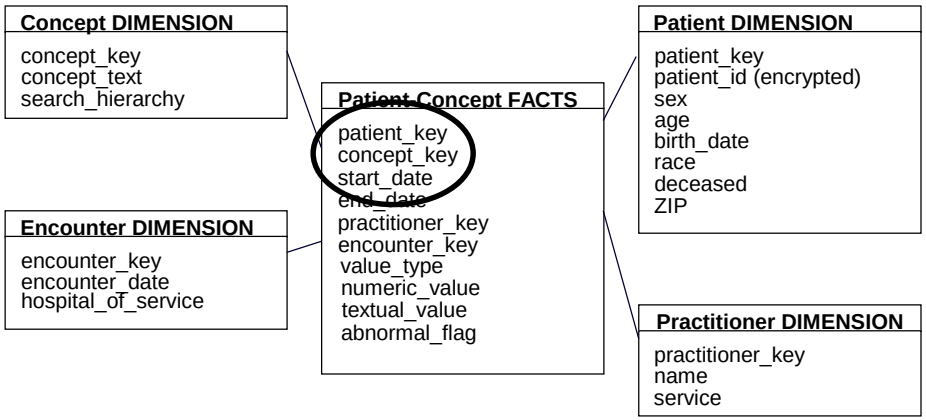
Patient Number	Gender	Age	Race	Parkinson's disease [Existence (Yes/No)]	Tobacco use disorder [Existence (Yes/No)]
2419	F	82	WHITE	Yes	Yes
2507	F	72	WHITE	Yes	Yes
3493	F	75	WHITE	Yes	Yes
6839	M	64	UNKNOWN	Yes	
18000	M	59	WHITE	Yes	
18336	M	82	WHITE	Yes	

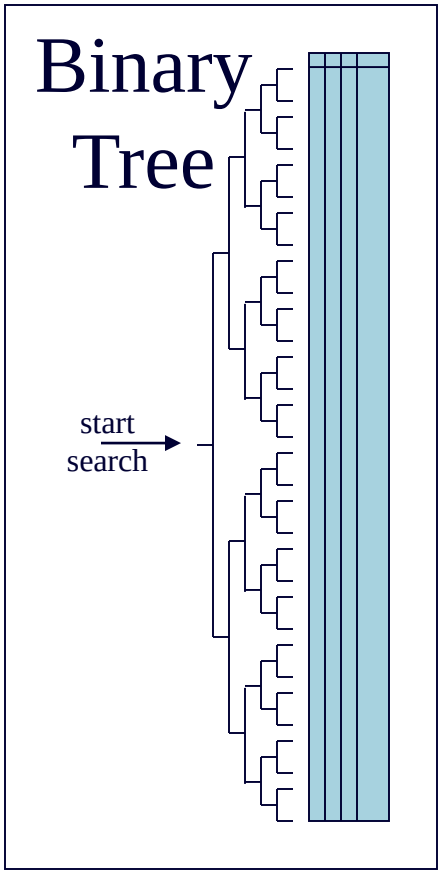
Cell Contains An Aggregated Value

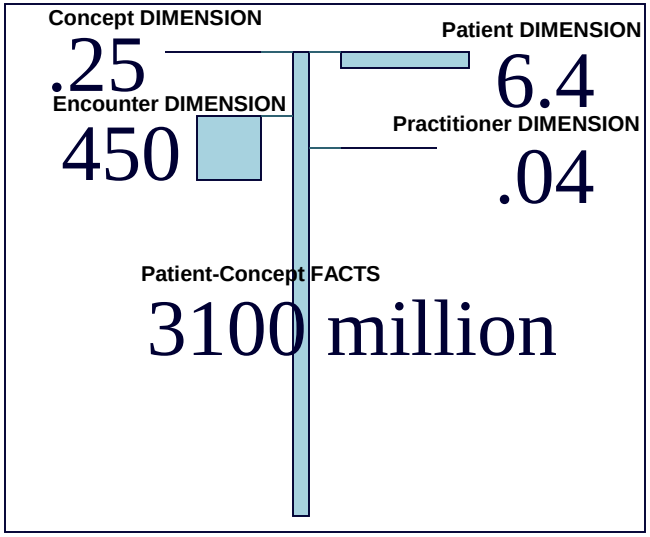
Existence (Yes/No)	Whether the patient has an observation of this concept. This is the default option.
Date (First)	Date of earliest observation
Date (Most Recent)	Date of the most recent observation
Count	Total number of observations
All Concepts (Names/Text)	All concept names are listed
Most Frequent Concept (Names/Text)	Most frequent concept name
All Concepts (Codes)	All concept codes are listed
Most Frequent Concept (Codes)	Most frequent concept code
Minimum Value	Minimum value of all numeric values observations
Maximum Value	Maximum value of all numeric values observations
Median Value	Median value of all numeric values observations
Average Value	Average value of all numeric values observations
Mode (Most Frequent Value)	Most frequent value for numeric, enum, blob observations
List of All Values	List of all values for numeric, enum, blob observations

Key for Success – Two types of tables

[2] Star Schema

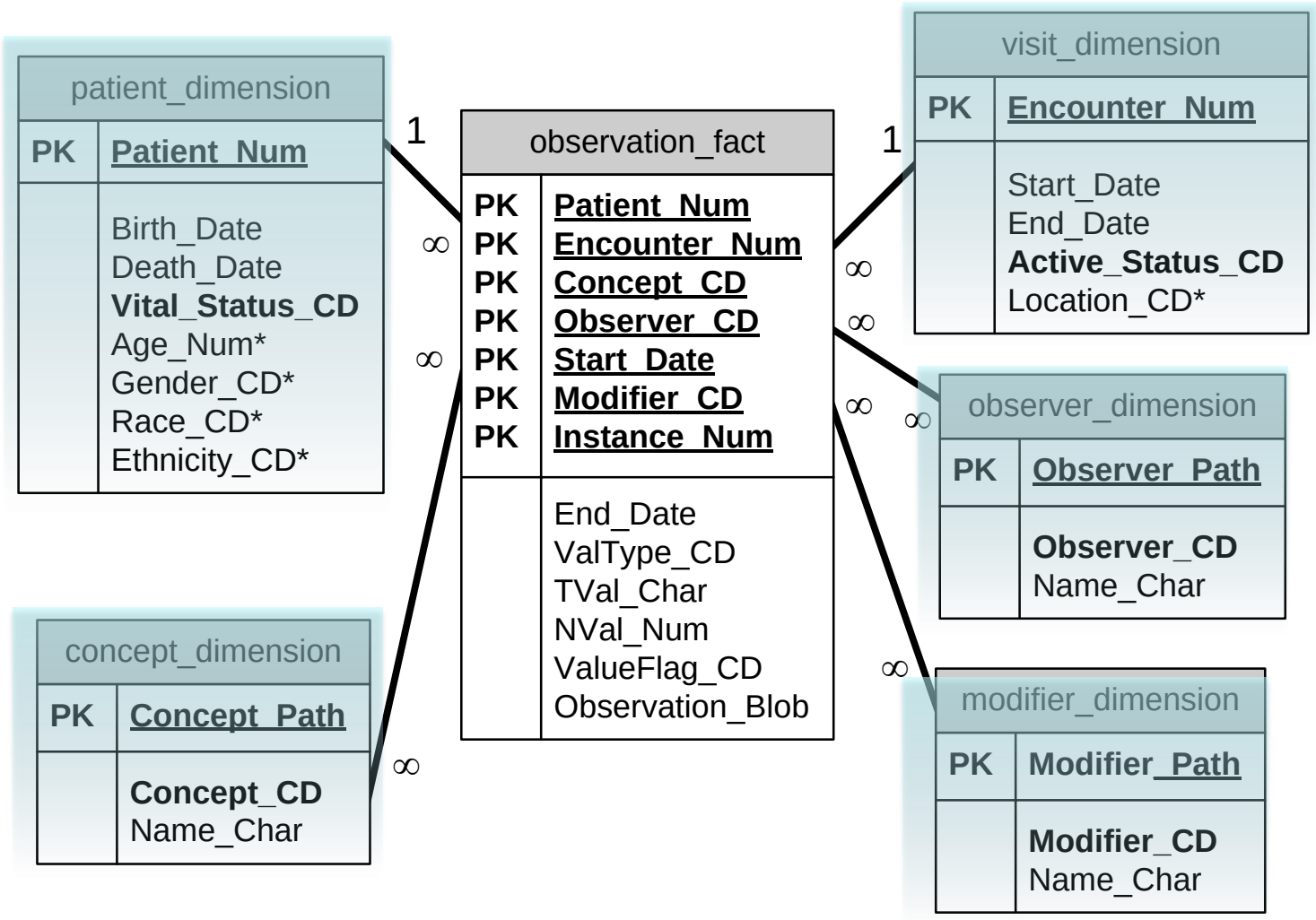




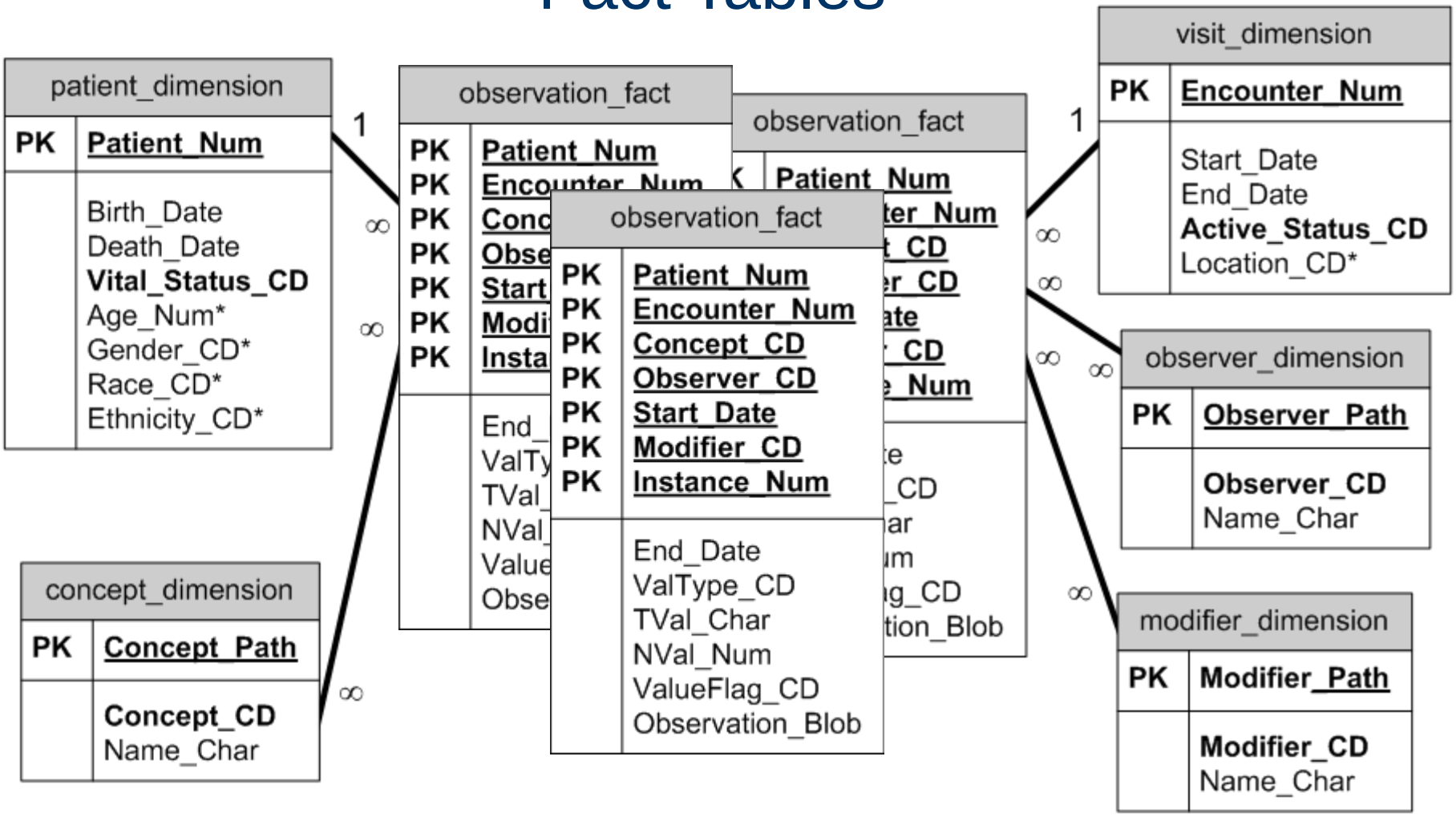


Theory of Kimball translated to Healthcare Data

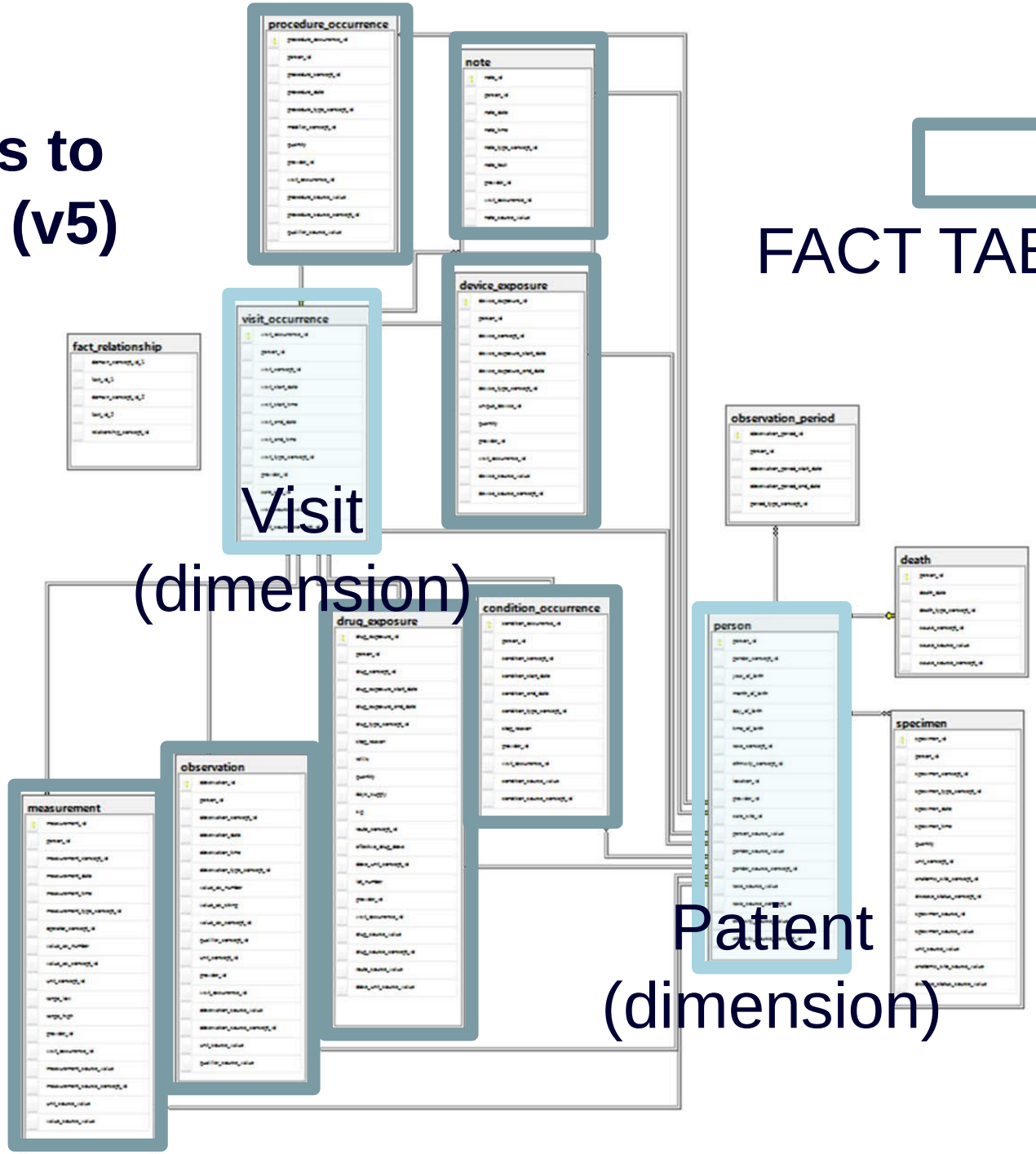
i2b2 Star Schema



i2b2 Star Schema can adapt to Multiple Fact Tables



Adapts to OMOP (v5)

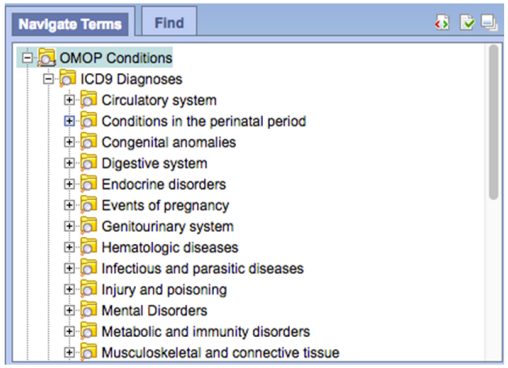


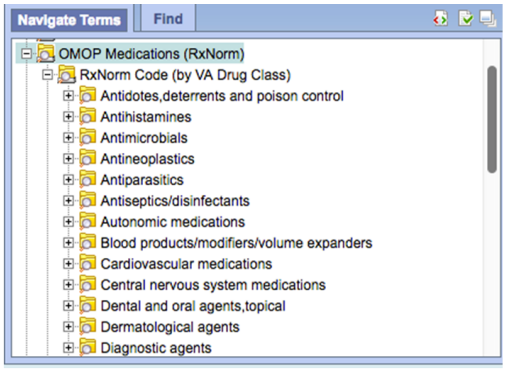
FACT TABLES

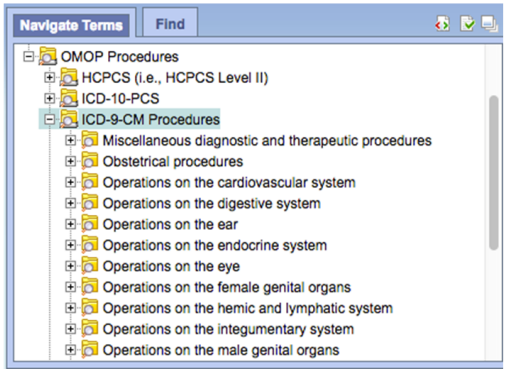
Visit (dimension)

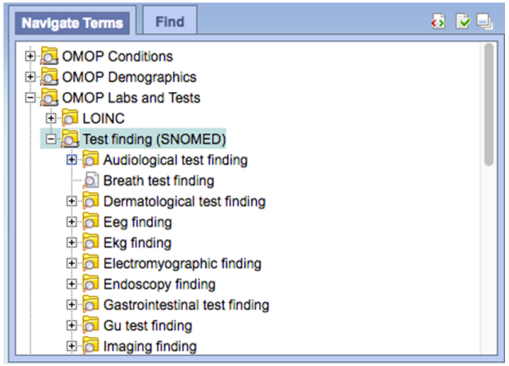
Patient (dimension)

Ontology Tables Need to be Created - Build ontology of OMOP standard concepts







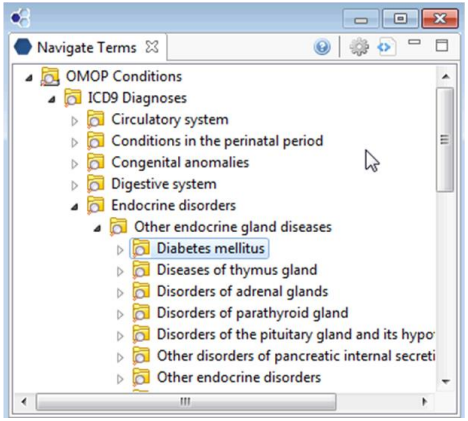


Ontologies covering the condition, procedures, drug, measurement and observation domains. All terms are mapped to standard concepts using OMOP's mapping tables

Use Ontology Tables to direct Queries to proper Fact Table view

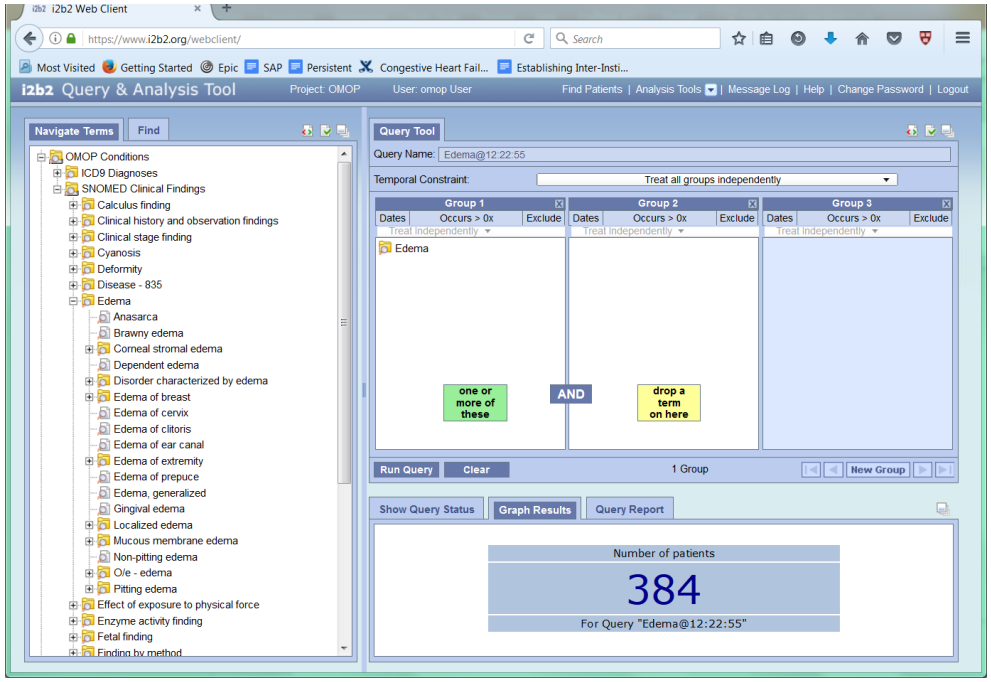
Prepend c_facttablecolumn with OMOP domain view and modify CRC to parse into ‘domain_view’ and ‘c_facttablecolumn’

c_name	c_facttablecolumn	c_tablename	c_columnname	c_operator	c_dimcode
Diabetes mellitus	condition_view.concept_cd	concept_dimension	concept_path	LIKE	\i2b2\Diagnoses\Endocrine disorders (240-259)\Other endocrine gland

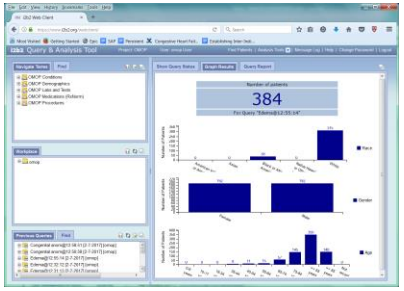



```
select patient_num from
condition_view where concept_cd
IN
(select
concept_cd from concept_dimension where
concept_path like '\i2b2\Diagnoses\Endocrine
disorders (240-259)\Other endocrine gland
diseases (250-259)\(250) Diabetes mellitus\%‘)
```

Linking OMOP into i2b2 software



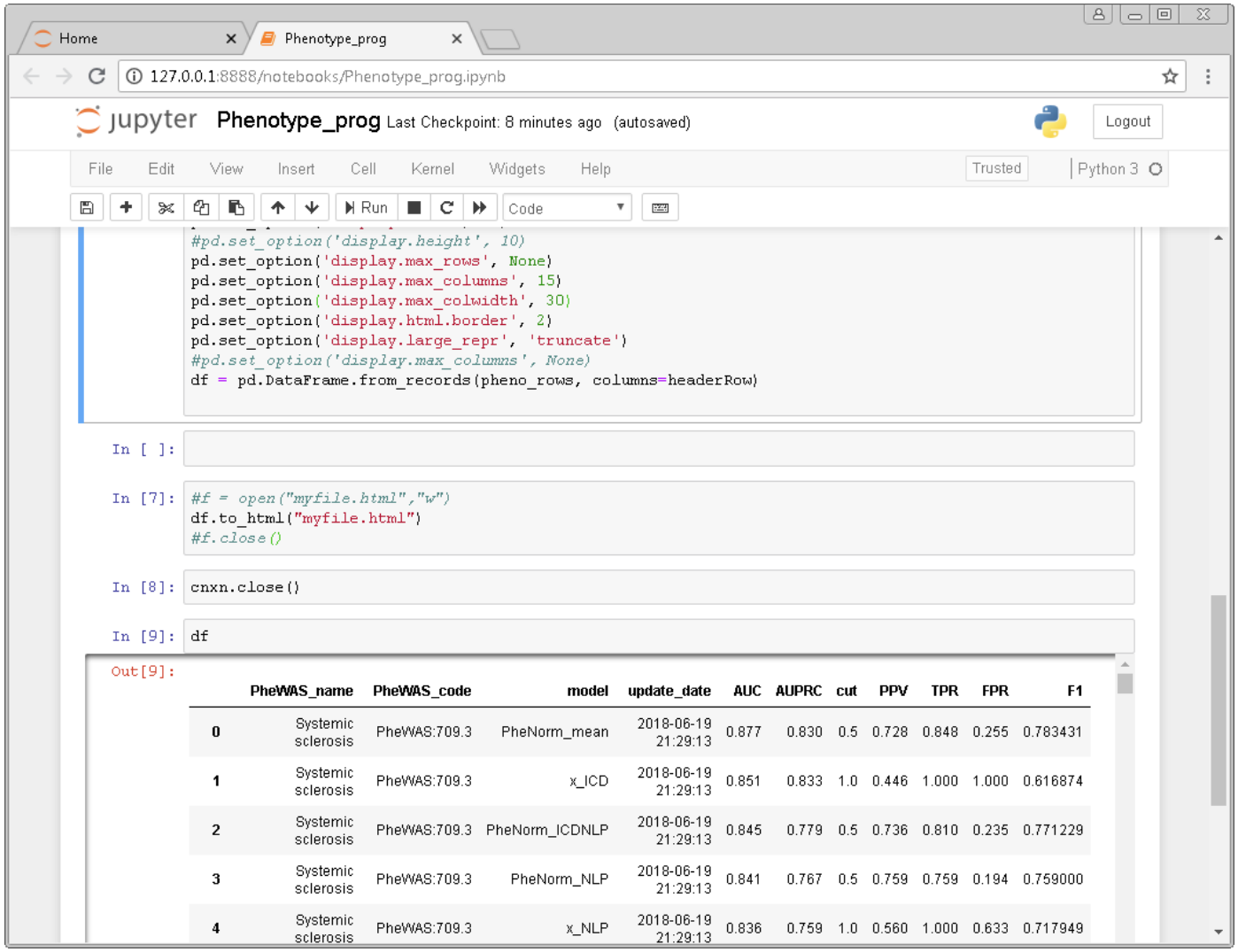
- **Medicare Claims Synthetic Public Use Files (SynPUFs) in OMOP v5 CDM is background data set**
- https://www.i2b2.org/webclient/
  - Username: omop
  - Password: demouser



Strategy is Converging ML/AI Starting Points

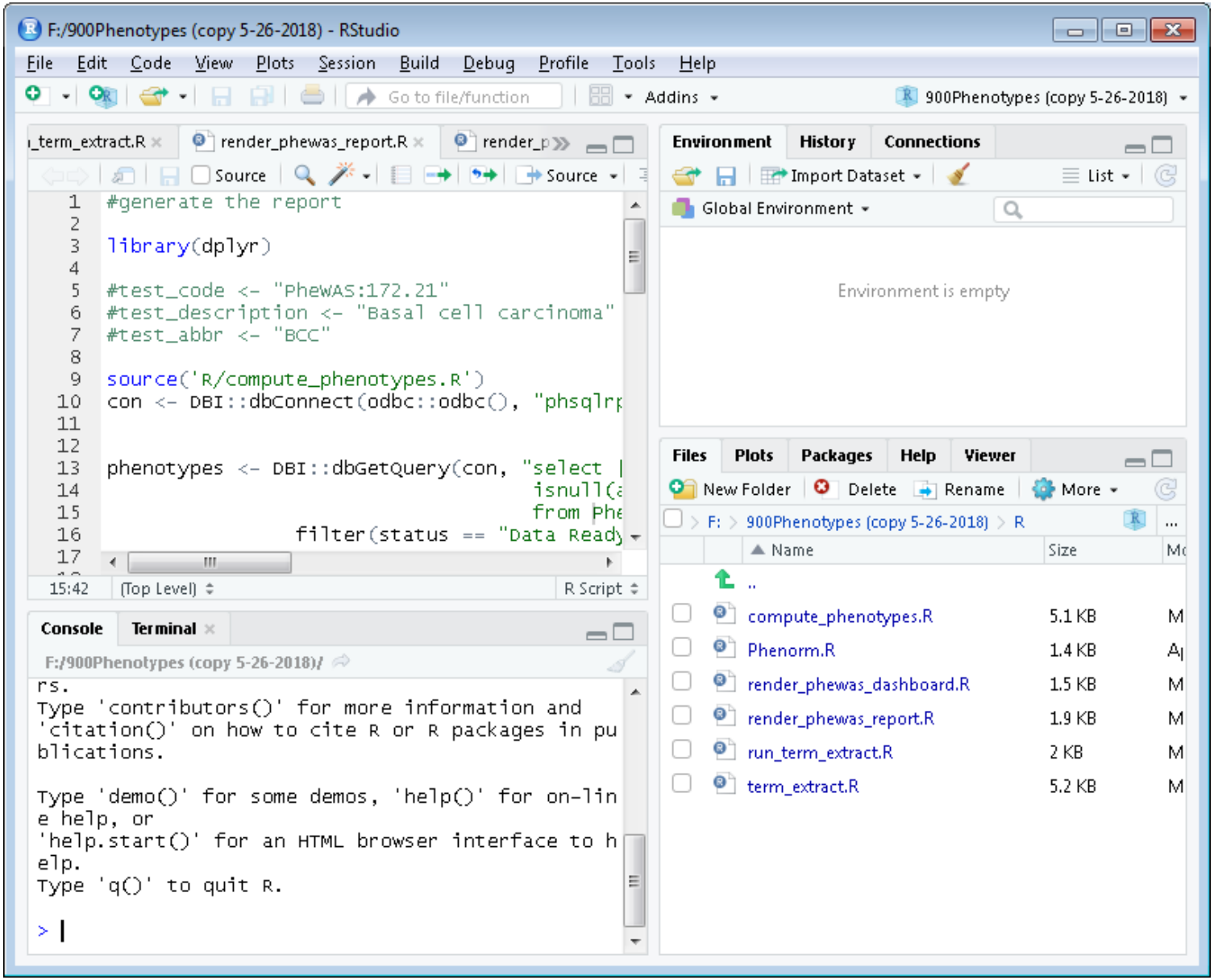
- Allow Learning Algorithms to grow from a common stating point
- In Data Enclave can accumulate libraries that operate from common data presentation
- Aggregate and Detail common presentations
- Transform other data models into these presentations so the look the same to computational algorithms
- Derived data can be stored back in relational structures at end of computation
- Serves as index for large file objects
- Derived data from large file objects goes back into the index
- Data marts of derived data can be created and published in i2b2 tranSMART DB

Jupyter Notebook

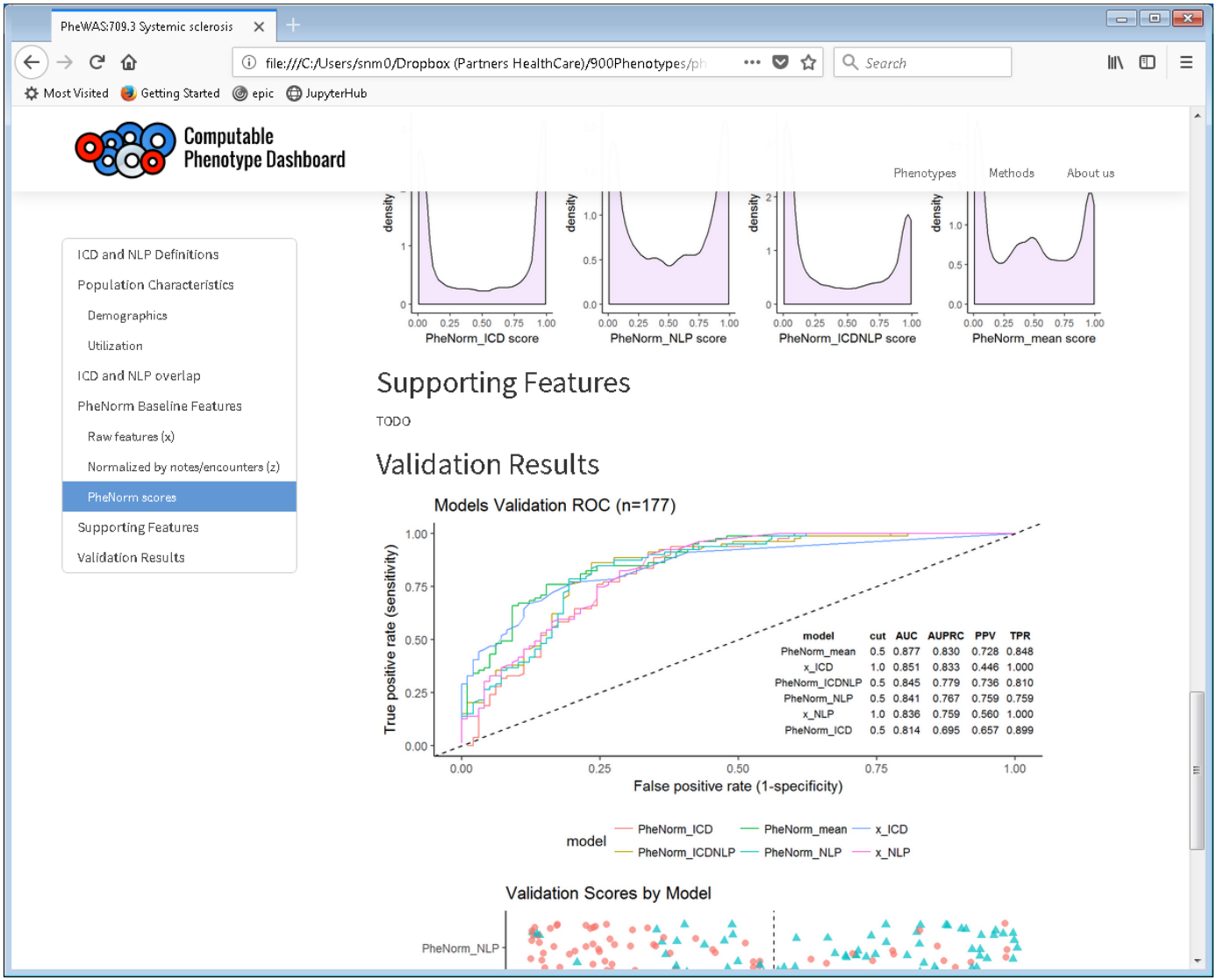


	PheWAS_name	PheWAS_code	model	update_date	AUC	AUPRC	cut	PPV	TPR	FPR	F1
0	Systemic sclerosis	PheWAS:709.3	PheNorm_mean	2018-06-19 21:29:13	0.877	0.830	0.5	0.728	0.848	0.255	0.783431
1	Systemic sclerosis	PheWAS:709.3	x_ICD	2018-06-19 21:29:13	0.851	0.833	1.0	0.446	1.000	1.000	0.616874
2	Systemic sclerosis	PheWAS:709.3	PheNorm_ICDNLP	2018-06-19 21:29:13	0.845	0.779	0.5	0.736	0.810	0.235	0.771229
3	Systemic sclerosis	PheWAS:709.3	PheNorm_NLP	2018-06-19 21:29:13	0.841	0.767	0.5	0.759	0.759	0.194	0.759000
4	Systemic sclerosis	PheWAS:709.3	x_NLP	2018-06-19 21:29:13	0.836	0.759	1.0	0.560	1.000	0.633	0.717949

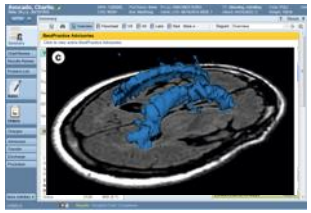
Python/R Environment



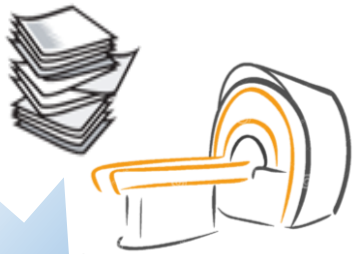
Python/R Reprots



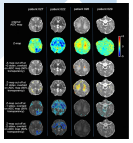
FHIR - GATHERING FULL SUPPLY CHAIN CONTROL



CLINICAL ENVIRONMENT



SMART Apps



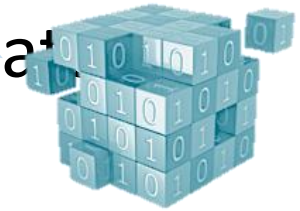
Clinical Data

INTEGRATION ENVIRONMENT

BIG DATA COMMONS

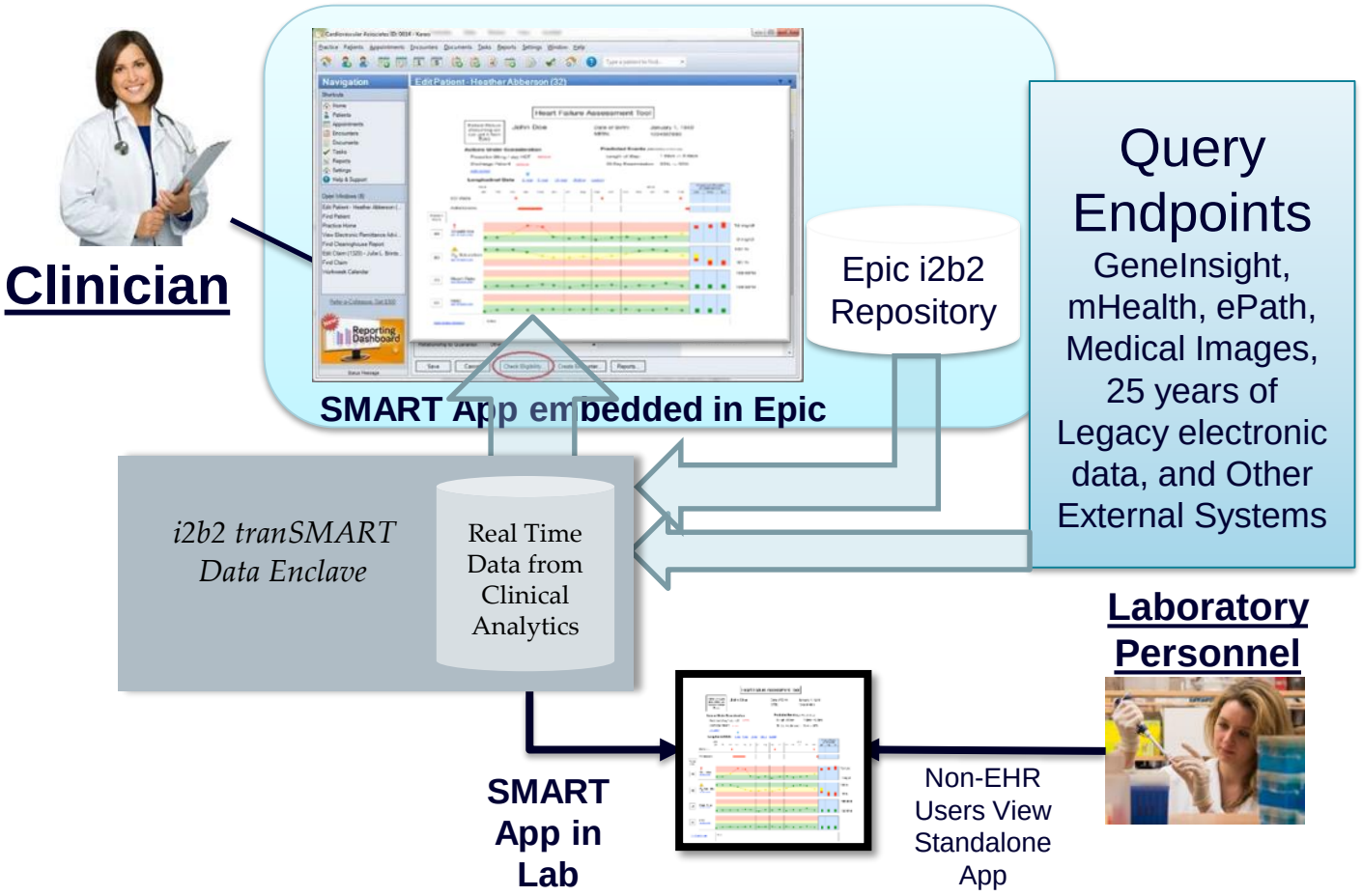
AI Applications

Structured Data

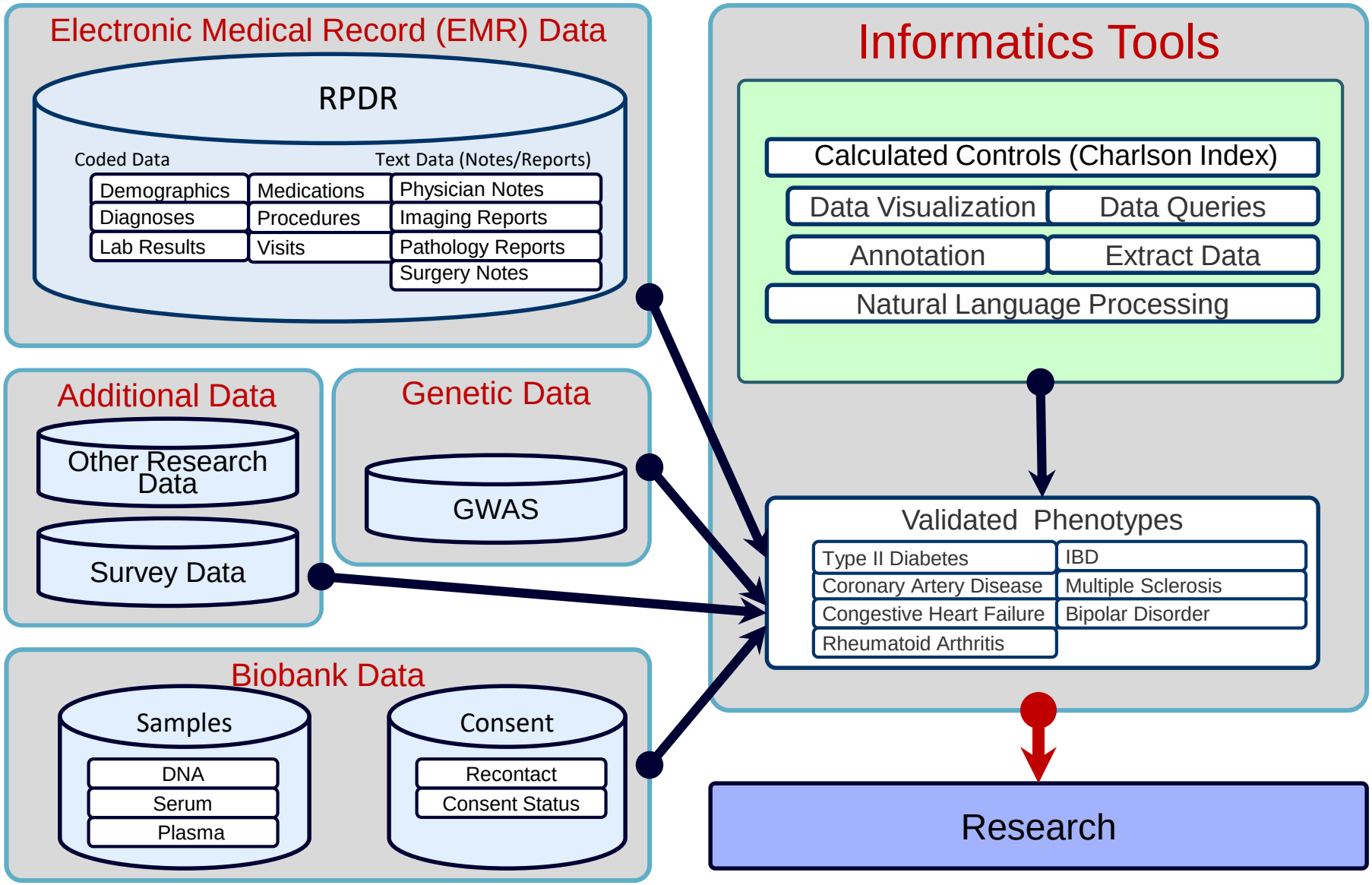


COMPUTATIONAL ENVIRONMENT

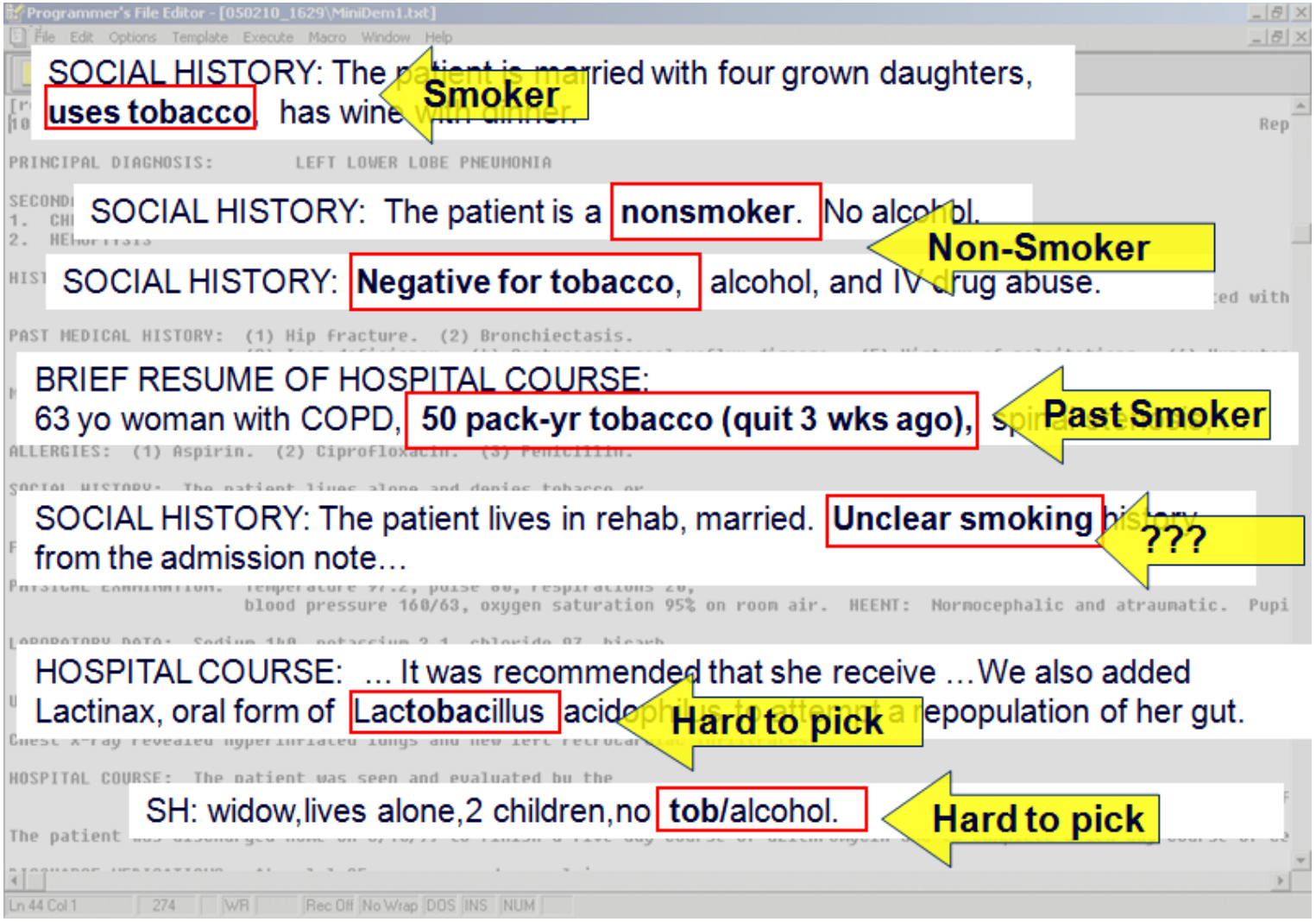
Data Enclave into Clinical Care



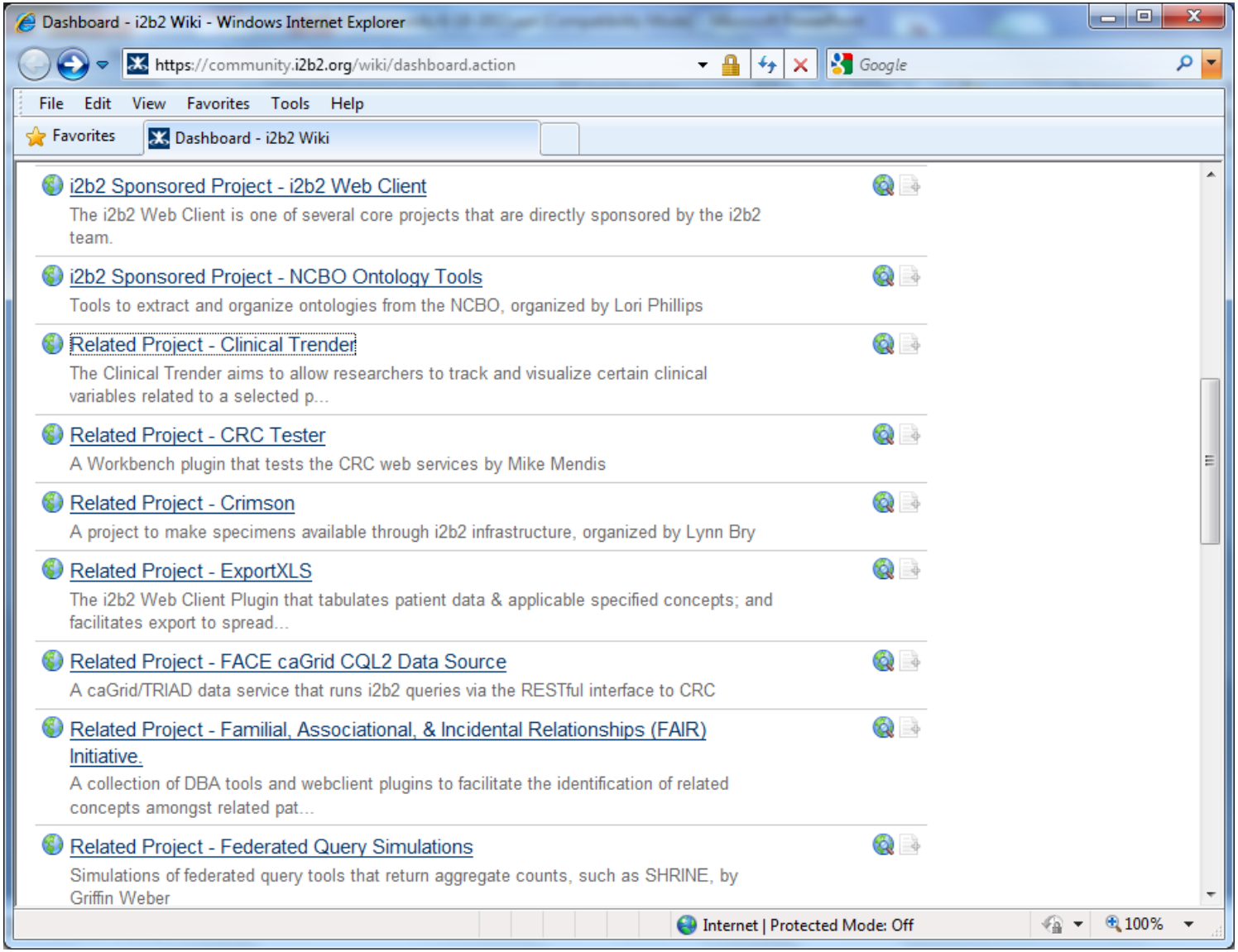
Data Integration in Data Enclave



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



I2b2 Community Software distributed as open source



Psychological Medicine (2012), **42**, 41–50. © Cambridge University Press 2011
doi:10.1017/S0033291711000997

ORIGINAL ARTICLE

Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model

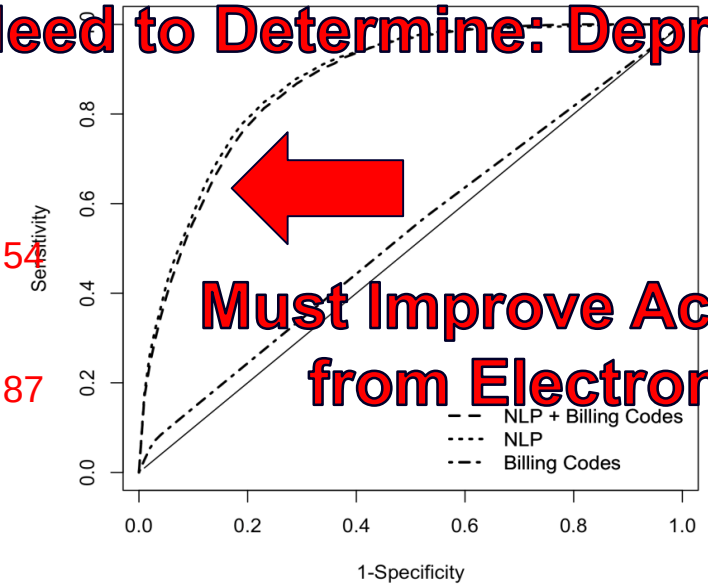
R. H. Perlis[1,2*], D. V. Iosifescu[1,3], V. M. Castro[4], S. N. Murphy[5], V. S. Gainer[4], J. Minnier[6], T. Cai[6], S. Goryachev[4], Q. Zeng[7], P. J. Gallagher[2], M. Fava[1], J. B. Weilburg[1], S. E. Churchill[8], I. S. Kohane[9] and J. W. Smoller[2]

Use Phenotyping Algorithms to define cohorts of treatment-resistant and treatment-responsive depression

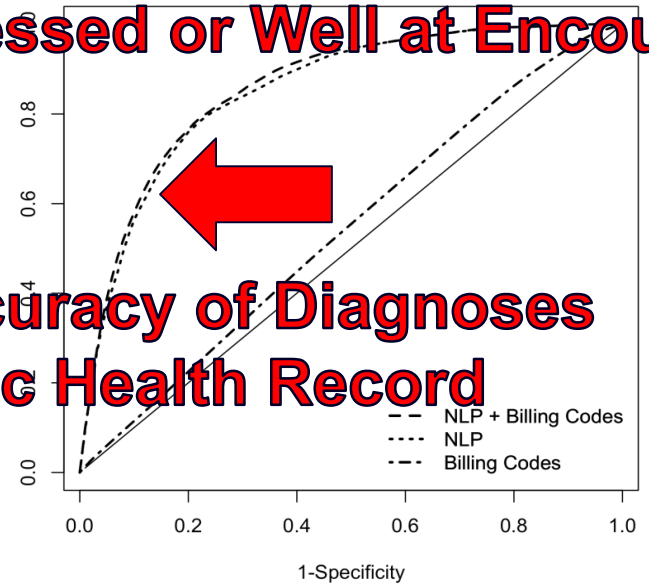
Need to Determine: Depressed or Well at Encounter

Initially:
AUC = 0.54

Finally:
AUC = 0.87



Must Improve Accuracy of Diagnoses from Electronic Health Record



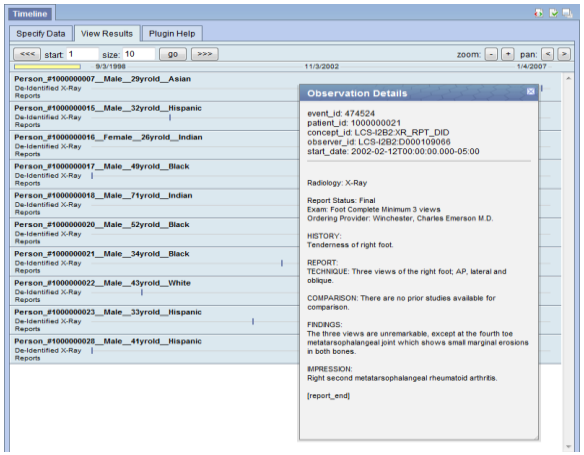
Initially:
AUC = 0.55

Finally:
AUC = 0.86

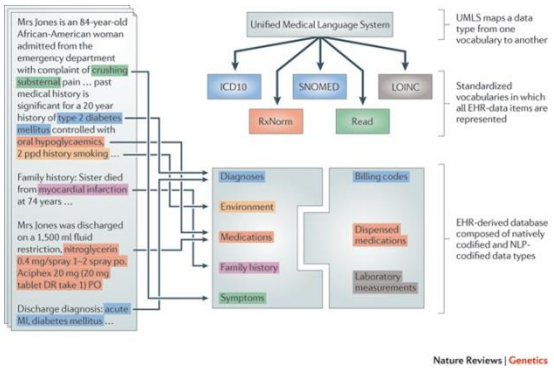
Clinical Status	Model	Specificity	Sensitivity	Precision	AUC
Depressed	Billing Codes	0.95	0.09 (0.03)	0.57 (0.14)	0.54 (0.02)
Depressed	NLP	0.95	0.42 (0.05)	0.78 (0.02)	0.88 (0.02)
Depressed	NLP + Billing Codes	0.95	0.39 (0.06)	0.78 (0.02)	0.87 (0.02)
Well	Billing Codes	0.95	0.06 (0.02)	0.26 (0.27)	0.55 (0.03)
Well	NLP	0.95	0.37 (0.06)	0.86 (0.02)	0.85 (0.02)
Well	NLP + Billing Codes	0.95	0.39 (0.07)	0.85 (0.02)	0.86 (0.02)

Curating a Disease Algorithm

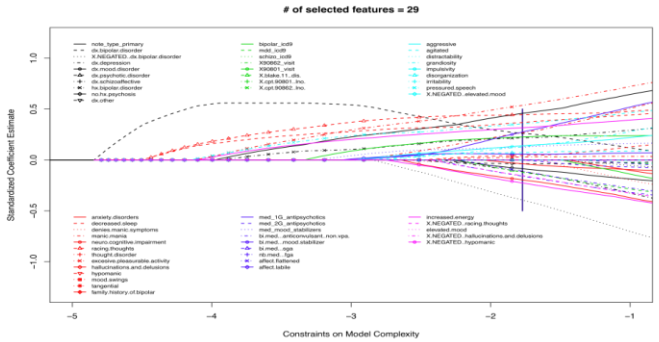
1. Create a gold standard training set.



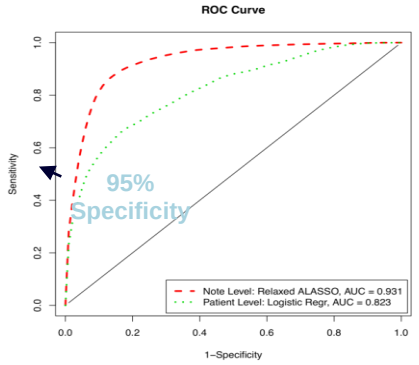
2. Create a comprehensive list of features from patient's electronic data that describe the disease of interest



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



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



Biobank Portal | Curated Diseases

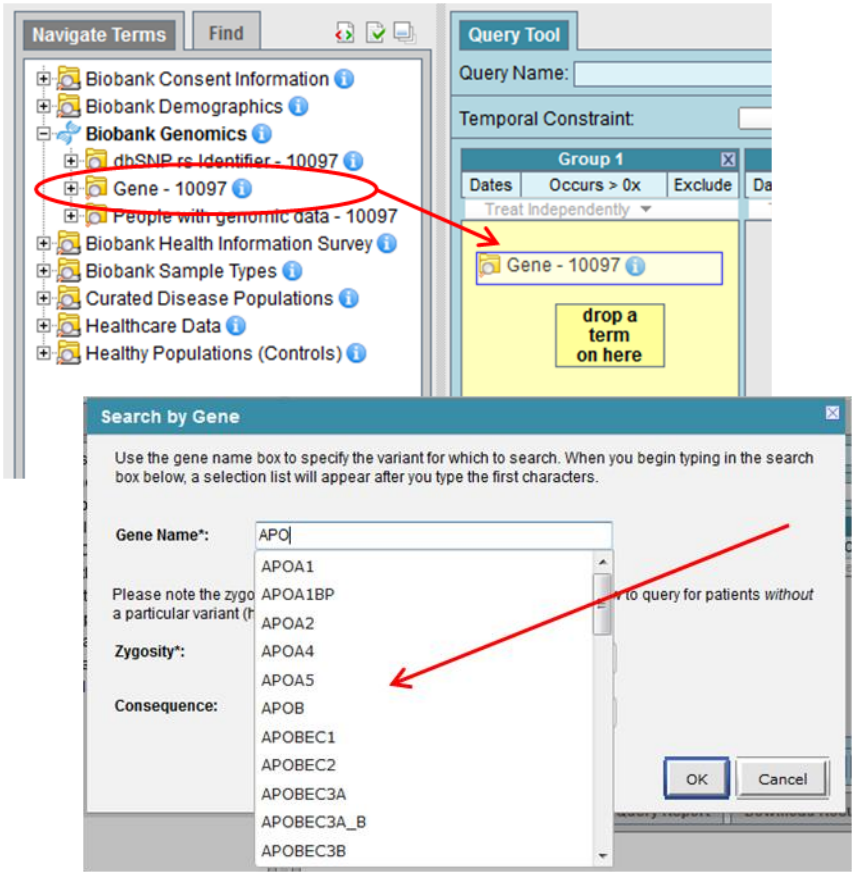
Validated Phenotype	Count*	Predictive Positive Value
Bipolar Disease	71	89%
Congestive Heart Failure	387	90%
Coronary Artery Disease	2,420	97%
Crohn's Disease	453	90%
Multiple Sclerosis	94	90%
Rheumatoid Arthritis	550	90%
Type 2 Diabetes Mellitus	1,887	97%
Ulcerative Colitis	330	90%

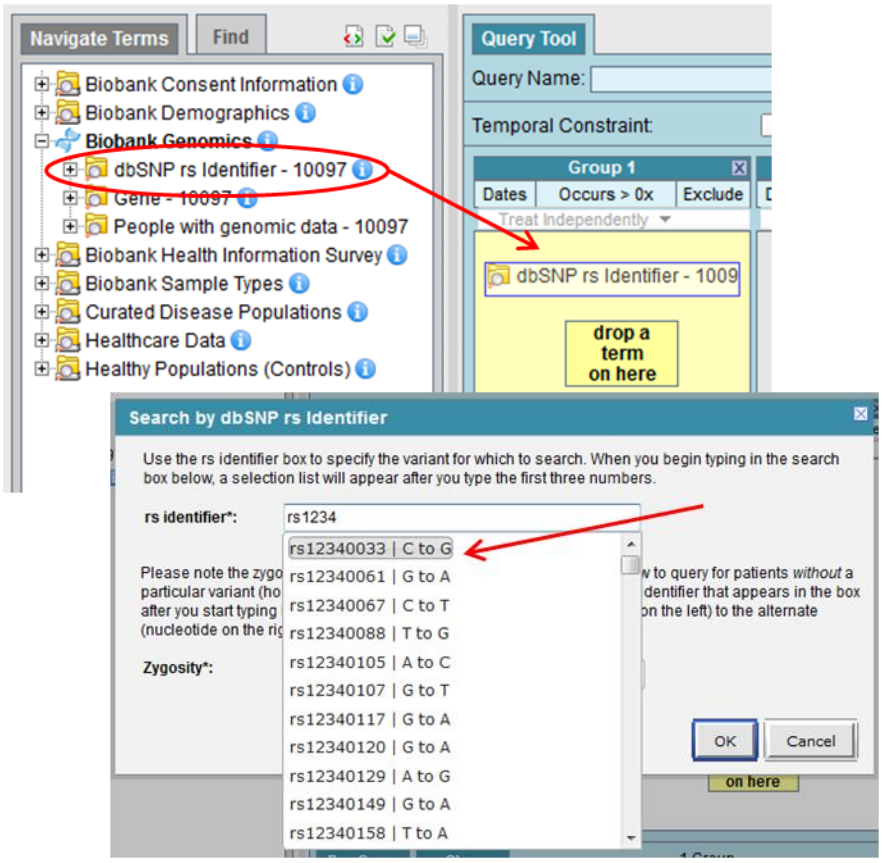
Healthy Controls based on Charlson Index	Count**
0 – 10-year survival probability is >98.3%	2,206
1 – 10-year survival probability is >95.87%	4,343
2 – 10-year survival probability is >90.15%	6,545

* Based on 15,880 patients

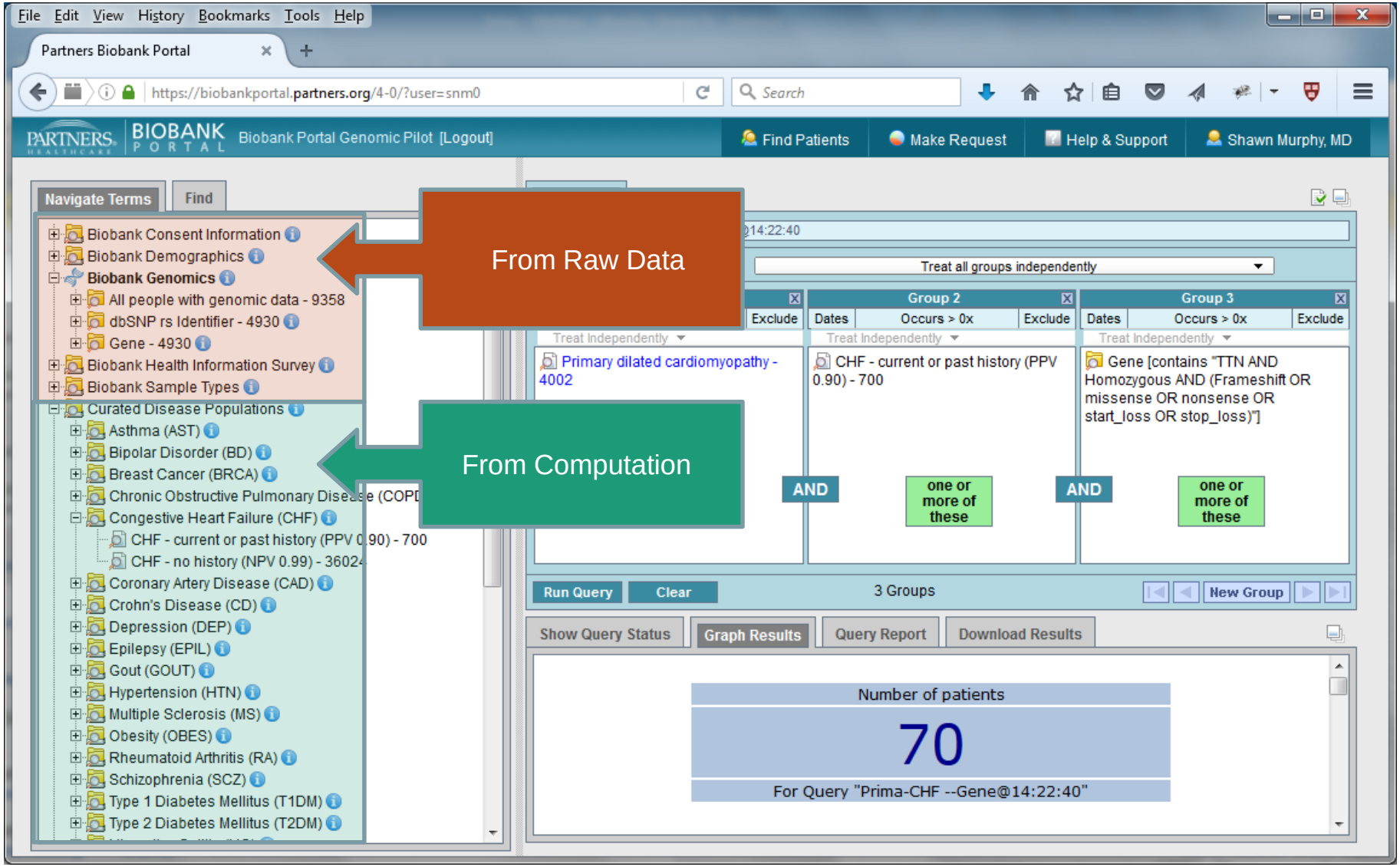
** Based on 21,300 patients

Genotype Data

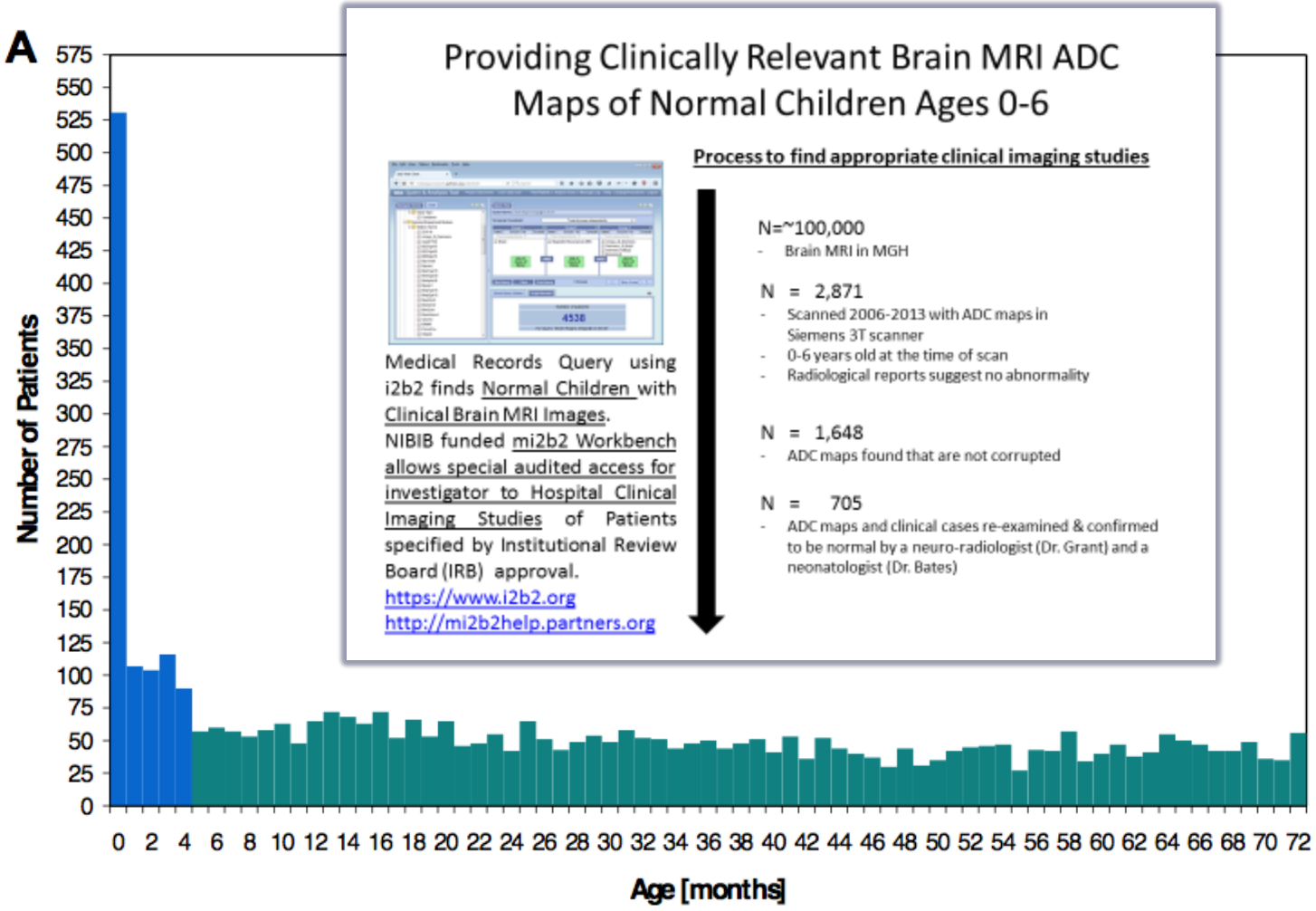




https://community.i2b2.org/wiki/display/IGD/Loading+Genomic+VCF+Files+into+i2b2


File Edit View History Bookmarks Tools Help
Partners Biobank Portal
https://biobankportal.partners.org/4-0/?user=snm0
Search
PARTNERS BIOBANK PORTAL
Biobank Portal Genomic Pilot [Logout]
Find Patients
Make Request
Help & Support
Shawn Murphy, MD
Navigate Terms
Find
Biobank Consent Information
Biobank Demographics
Biobank Genomics
All people with genomic data - 9358
dbSNP rs Identifier - 4930
Gene - 4930
Biobank Health Information Survey
Biobank Sample Types
Curated Disease Populations
Asthma (AST)
Bipolar Disorder (BD)
Breast Cancer (BRCA)
Chronic Obstructive Pulmonary Disease (COPD
Congestive Heart Failure (CHF)
CHF - current or past history (PPV 0.90) - 700
CHF - no history (NPV 0.99) - 3602
Coronary Artery Disease (CAD)
Crohn's Disease (CD)
Depression (DEP)
Epilepsy (EPIL)
Gout (GOUT)
Hypertension (HTN)
Multiple Sclerosis (MS)
Obesity (OBES)
Rheumatoid Arthritis (RA)
Schizophrenia (SCZ)
Type 1 Diabetes Mellitus (T1DM)
Type 2 Diabetes Mellitus (T2DM)
From Raw Data
From Computation
14:22:40
Treat all groups independently
Group 2
Group 3
Dates
Occurs > 0x
Exclude
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)"]
AND
one or more of these
AND
one or more of these
Run Query
Clear
3 Groups
New Group
Show Query Status
Graph Results
Query Report
Download Results
Number of patients
70
For Query "Prima-CHF --Gene@14:22:40"

Find Normal Brain MRI's of Children

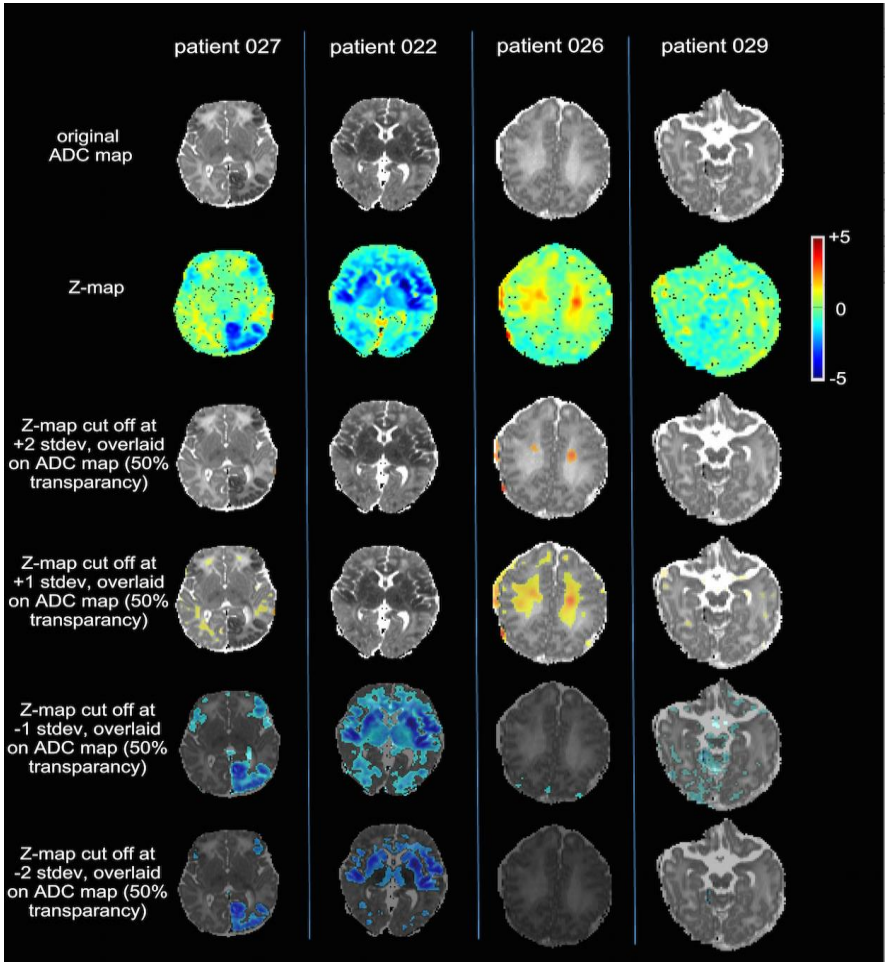


Number of patients who had a brain MRI scan at a particular age in months from 0 to 6 years (A) and in weeks from 0 to 4 months (B)

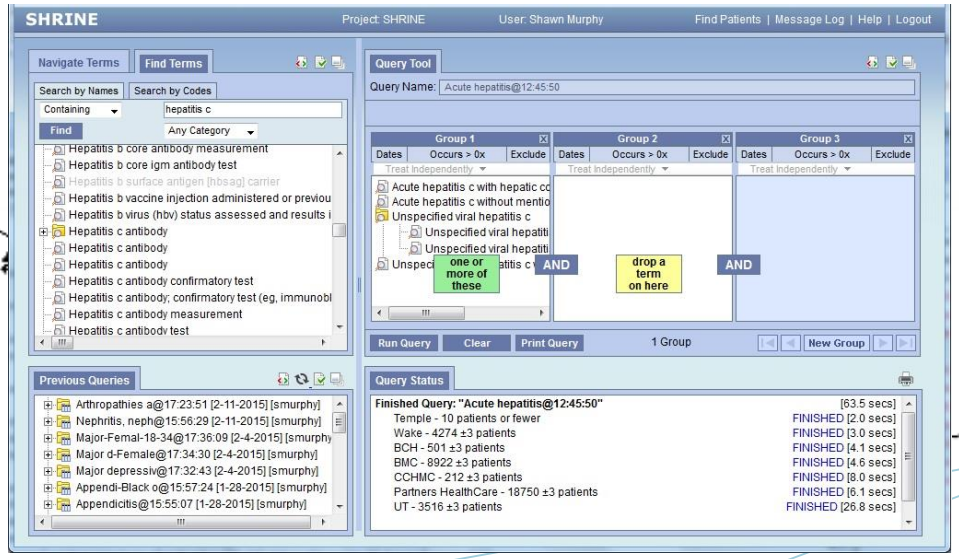
Atlases provide a visual guide for Radiology Decision Support, such as determining Perinatal Hypoxic Ischemic Encephalopathy

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

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



Federated Queries



Partners HealthCare System

Boston Children's Hospital

BIDMC

Boston Health Net (BMC and Community Health Centers)

Columbia U. Medical Center and New York Presbyterian Hospital

University of California, Davis

Washington University in St. Louis

Wake Forest Baptist Medical Center

Morehouse/Grady/RCMI

U Texas Health Science Center/Houston

I2b2 Implementations

CTSA's

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

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

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

HMOs:

- Group Health Cooperative
- Kaiser Permanente

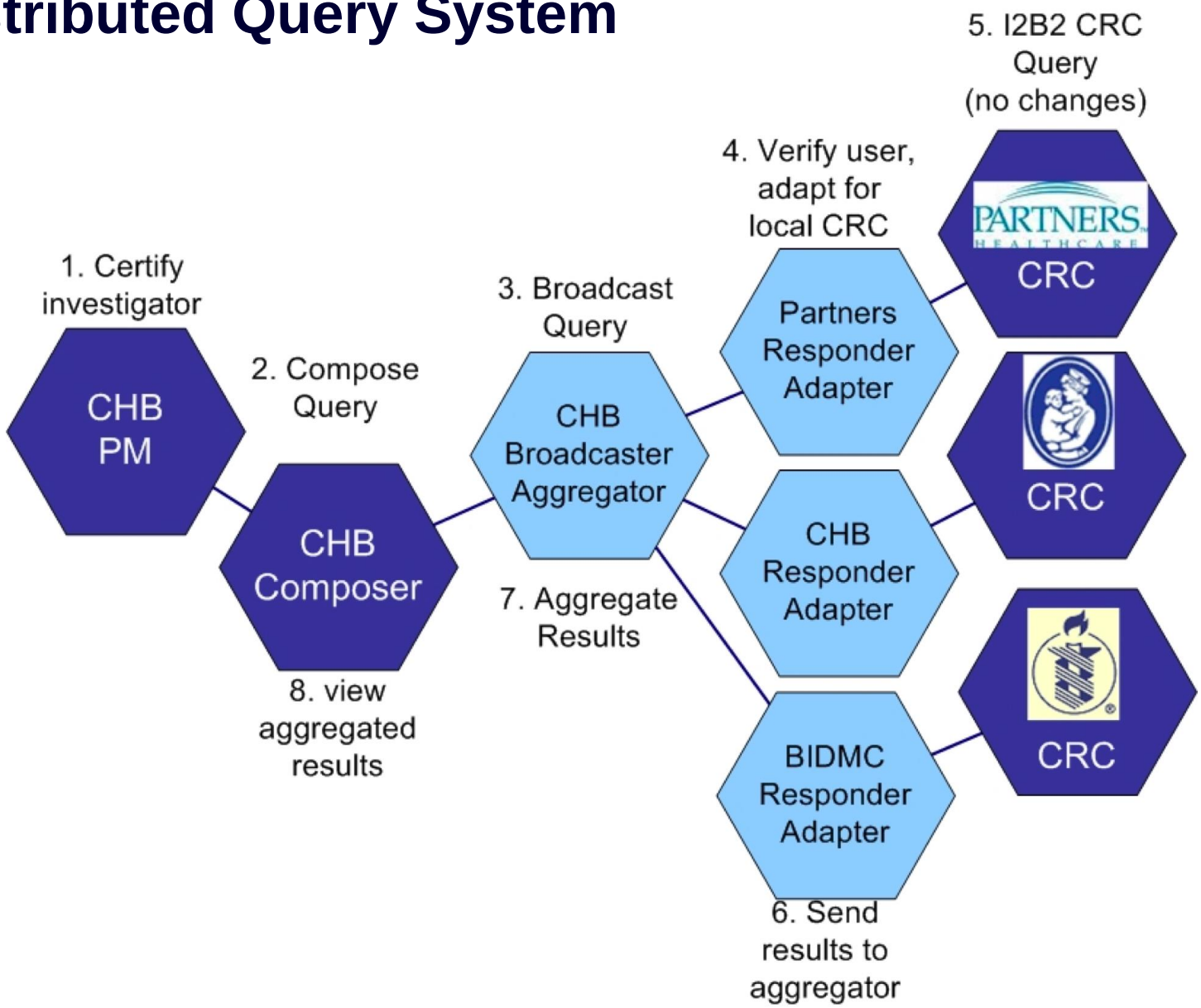
International:

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

Companies:

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

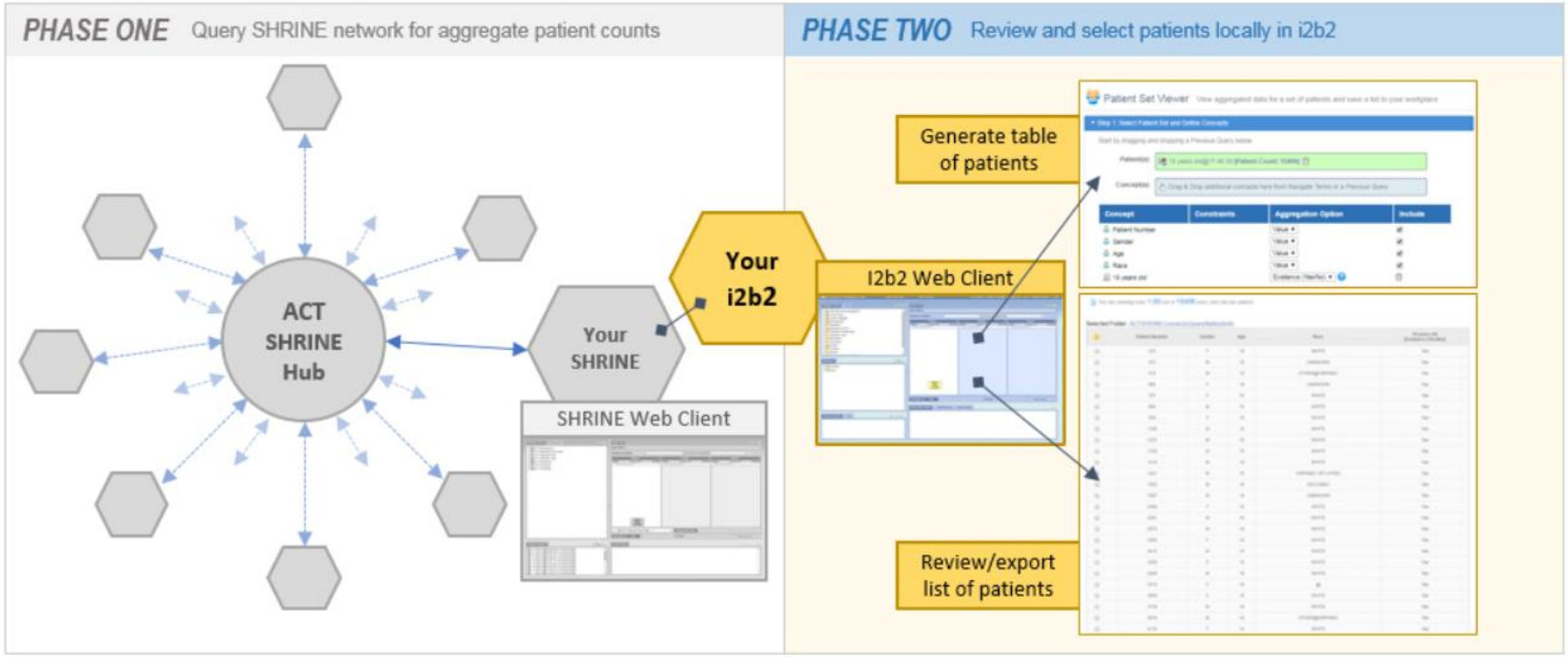
Distributed Query System



ACT 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

Accrual for Clinical Trials (ACT) - Workflow



eMERGE network

ELECTRONIC MEDICAL RECORDS AND GENOMICS

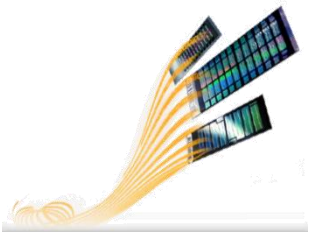
2500

25,000 Network-wide

Sequencing 100 high-priority genes

PARTNERS HEALTHCARE BIOBANK

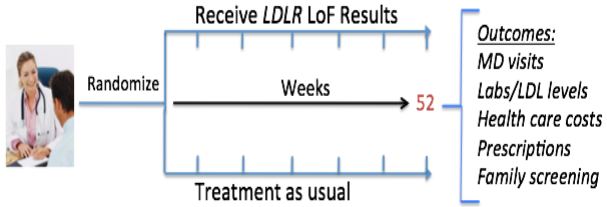
25,000



Genotyping

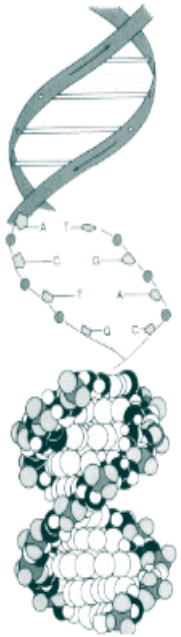
Discovery

Penetrance and Pleiotropy



Tribute to…

- Jeff Klann
- Kavi Wagholikar
- Lori Phillips
- Isaac Kohane
- Kenneth Mandl
- Joshua Mandel
- Griffin Weber
- Paul Avillach
- Christopher Herrick
- Alyssa Goodson
- Michael Mendis
- Vivian Gainer
- Victor Castro
- Nich Wattanasin
- Wayne Chan
- David Wang
- Andrew Cagan
- Bhaswati Ghosh
- Retta Metta
- Adam Landman
- Willian Gordon



I2b2, SHRINE, and SMART Information and Software on the Web

i2b2 Homepage (https://www.i2b2.org)

i2b2 Software (https://www.i2b2.org/software)

i2b2 Community Site (https://community.i2b2.org)

SMART Platforms Homepage (http://smarthealthit.org)

THANK YOU