

I2b2 User Guide

October 5, 2015

Prepared By: Nathaniel Poznanski-Ring, Data Analyst Regenstrief Institute

OVERVIEW

Introduction

Task 1: Patients in the ER at Eskenazi

- Interface/Drag and Drop

- Drilling down

- Modify the query to focus on ER

- Modify the query to provide demographics

Task 2: Patients with normal systolic blood pressure since 2010

- Finding labs in the *find* section

- Selecting the appropriate test

- Modifying the query to look for BP of between 120 and 139

- Interacting with the time function

- Breaking results down by vital status

Task 3: Patients taking acetaminophen AND/OR antihistamines

- Finding the method for identifying medications

- Selecting classes or sub-classes

- Moving data to the AND/OR conditions

Task 4: Patients with Pain

- Finding pain patients in dx and complaints

- Combining results with acetaminophen

Using IUH data

Task 1: Diabetes

Task 2: Long QT

Task 3: Testicular Cancer

Introduction

i2b2 is a web-based clinical research exploration tool, that enables on researchers to refine and define a patient or concept population quickly using a drag-and-drop interface. The software is widely used at other institutions and is based on a common framework. However, the Regenstrief deployment offers several unique advantages specifically in terms of depth and breadth of information. Unlike other deployments of i2b2, the Regenstrief deployment allows researchers the opportunity explore clinical data for Indiana University Health and Eskenazi health. At this time the two institutions are not searched simultaneously. However, this need will be met with an upcoming service called, ACT.

To gain access to the system, researchers will require creating an account with Indiana University: CTSI: <https://beta.indianactsi.org/regenstrief/i2b2-account-request>. Once you do this, you will create a JIRA ticket through the CTSI portal. Once you have completed the form you should receive an email letting you know that the ticket has been submitted. Once the account has been created, i2b2 can be accessed at <https://i2b2ctsi.regenstrief.org/>. From there you will log into system using your home institution credentials.

There are several caveats you should be aware of while using i2b2: the deidentified nature of the data, and the structure of the data; different sources of data can be different both qualitatively and quantitatively; . First, any data with less than 10 persons has been removed from the interface. This is to remain HIPPA compliant and generally does not affect the quality of a query. Secondly, the data is modeled after the Regenstrief concept tree. This means that individuals familiar with Careweb or G3 will recognize many of the hierarchies, whereas individuals who have not spent much time with Careweb or G3 may need to spend some time to familiarize themselves with the structure. Fortunately, for either case we'll quickly show you how to kick-start that learning curve.

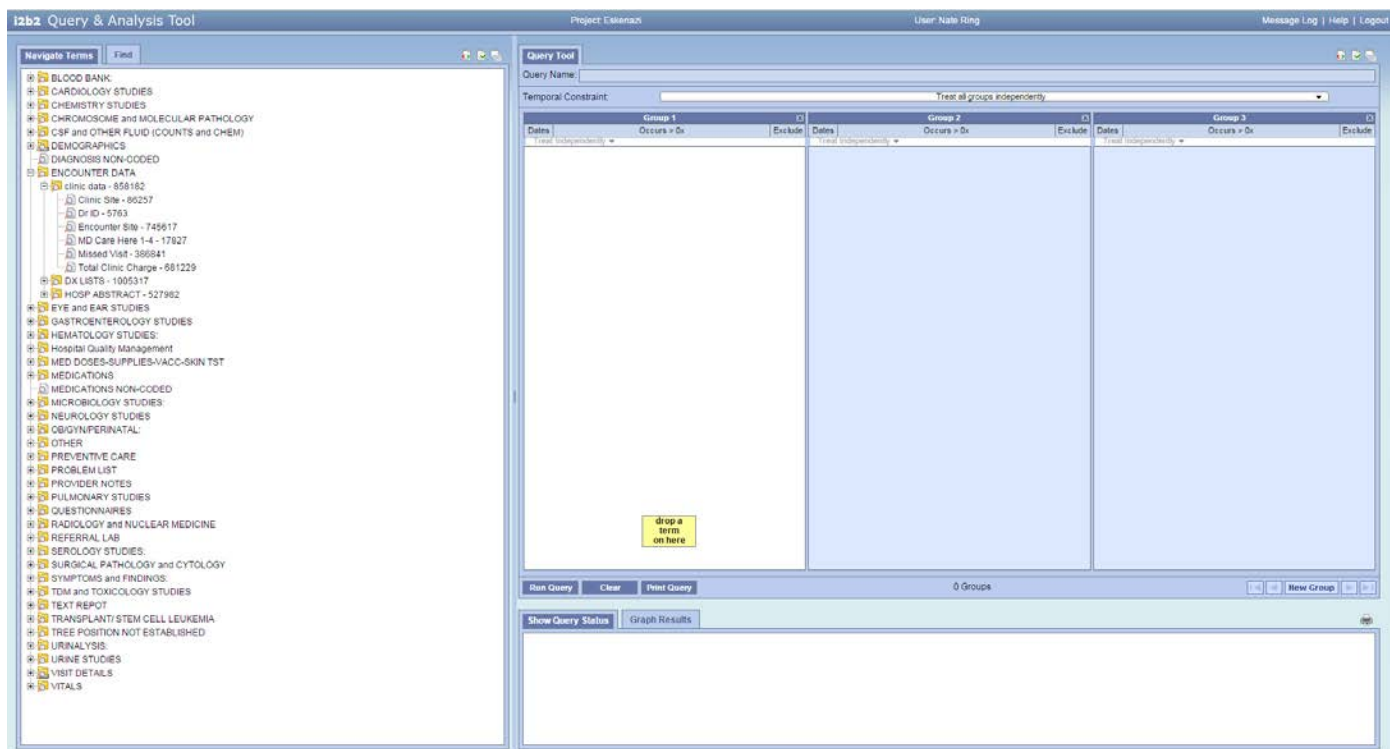
Let's explore together four quick hypothetical searches in Eskenazi: first patients seen at a site, second finding labs, third patients discharged with medications, and finally patients with a diagnosis or complaint.

Then we will explore some queries in the IUH silo: Metabolic diseases, Long QT, and Testicular cancer.

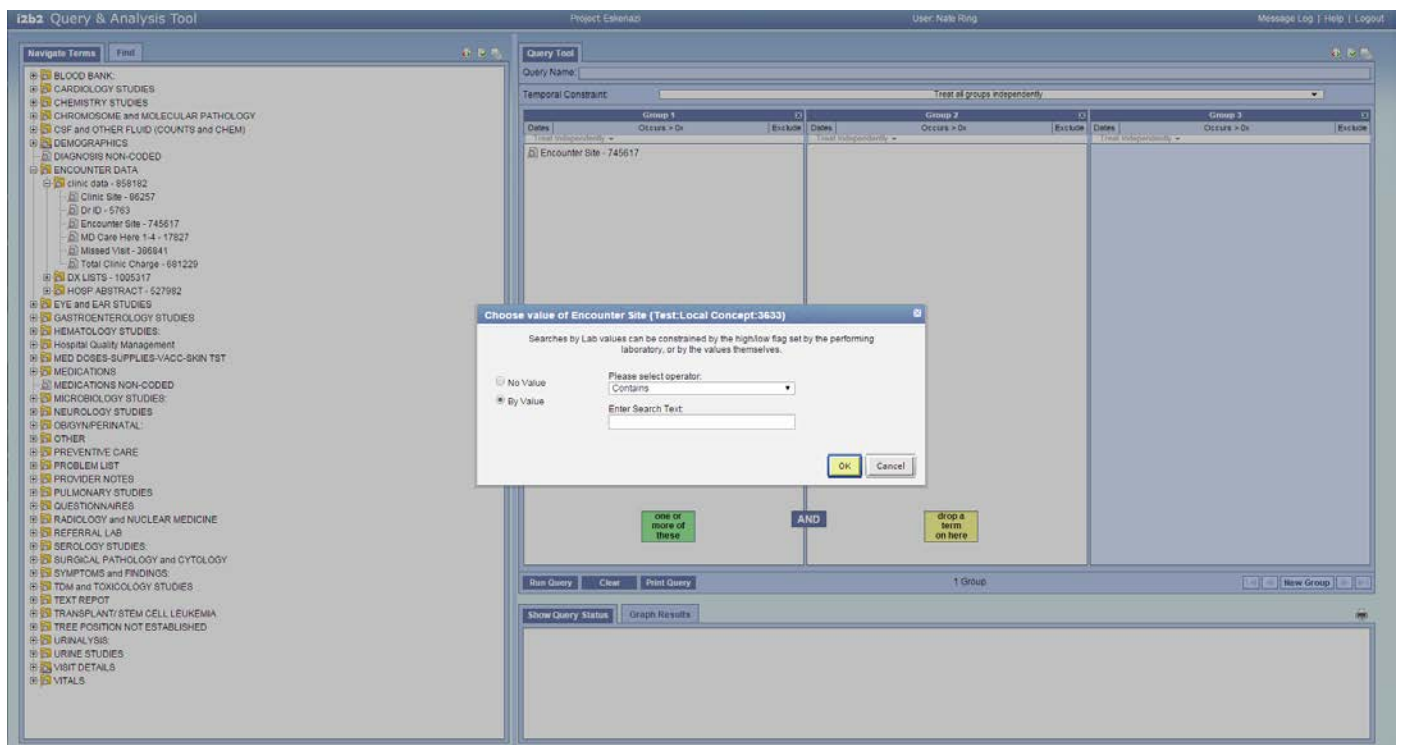
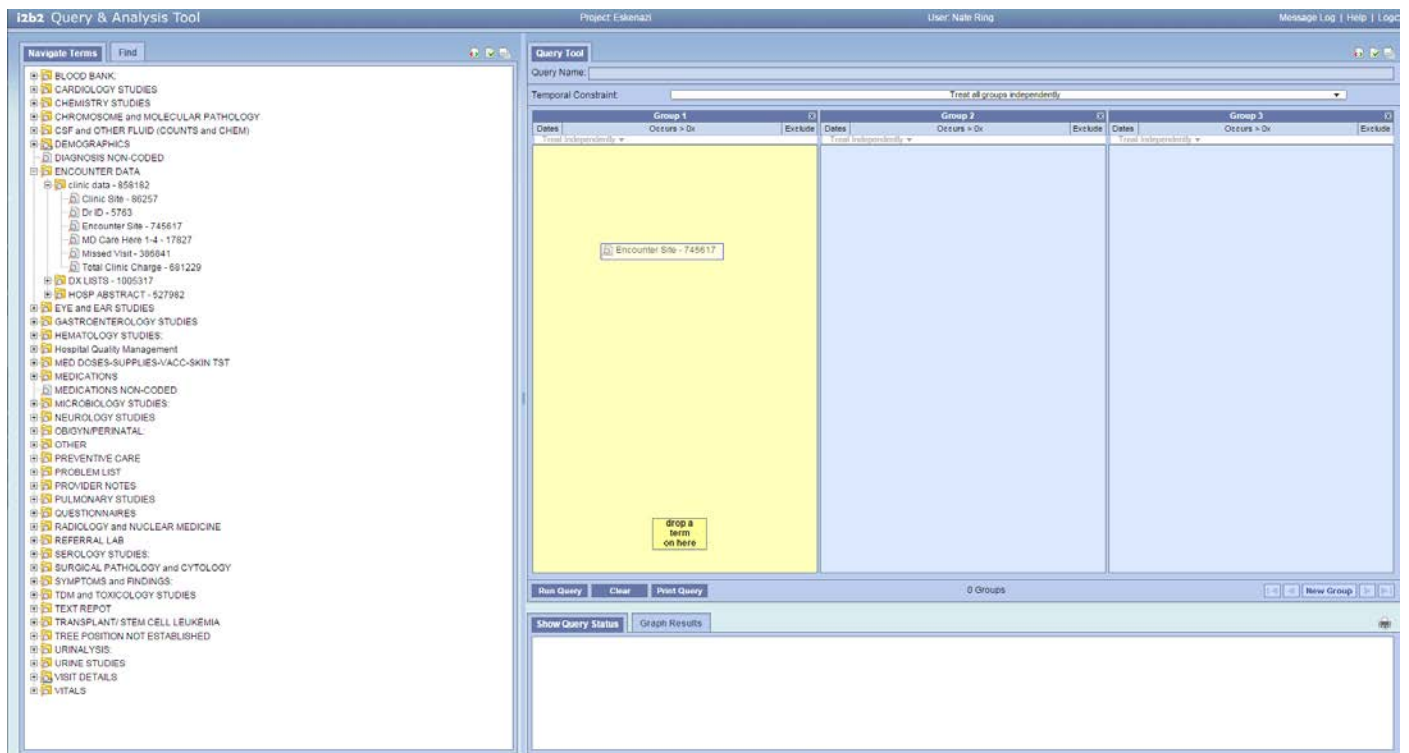
Task 1: Encounters

Let's see what kind of patients we can find whom have gone to the ER at Eskenazi. We'll then refine the query to see the number by gender, then by age.

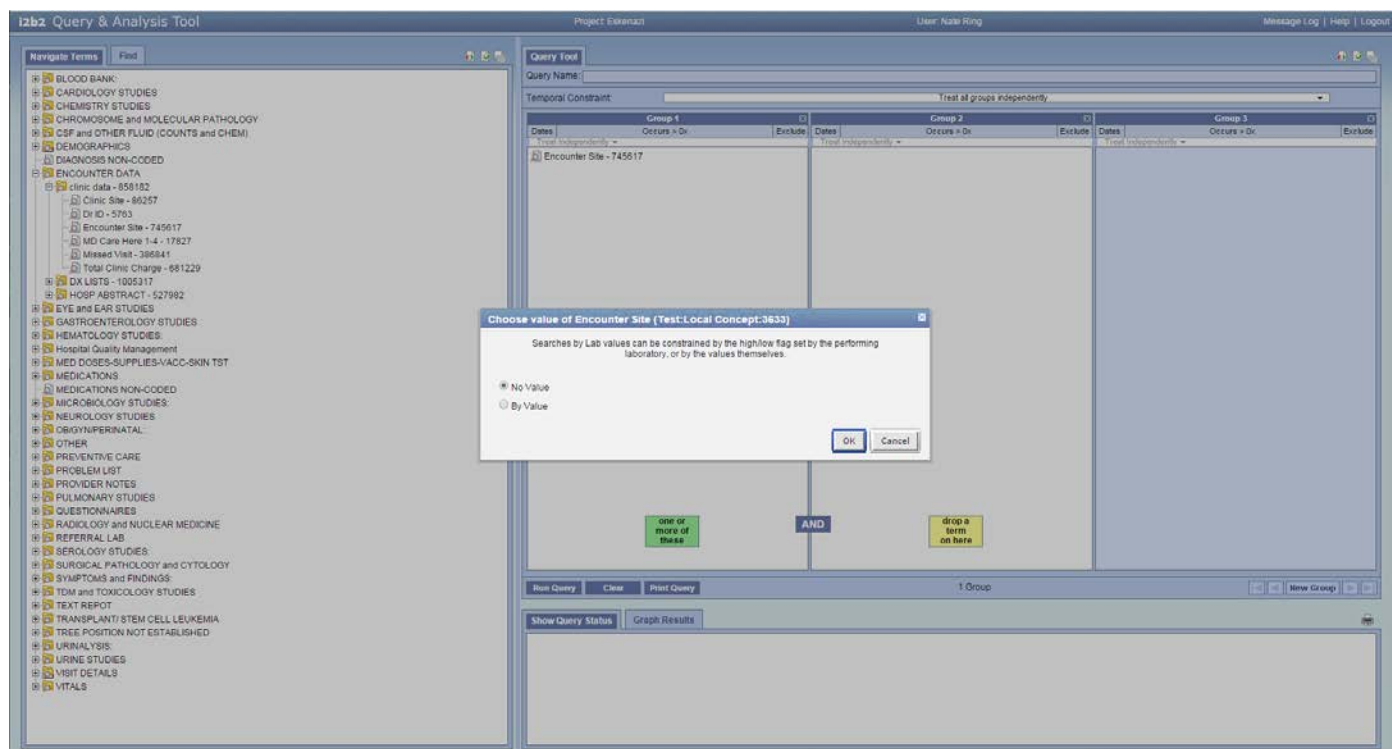
The first thing you should do is go to the ENCOUNTER DATA directory. Click on the plus to expand the tree. There you'll see all of the concepts related to encounters, such as clinic data, DX lists (diagnoses), and hospital abstracts. We are interested in the clinic data, so expand the *clinic data* folder. Your screen should look like the image below.



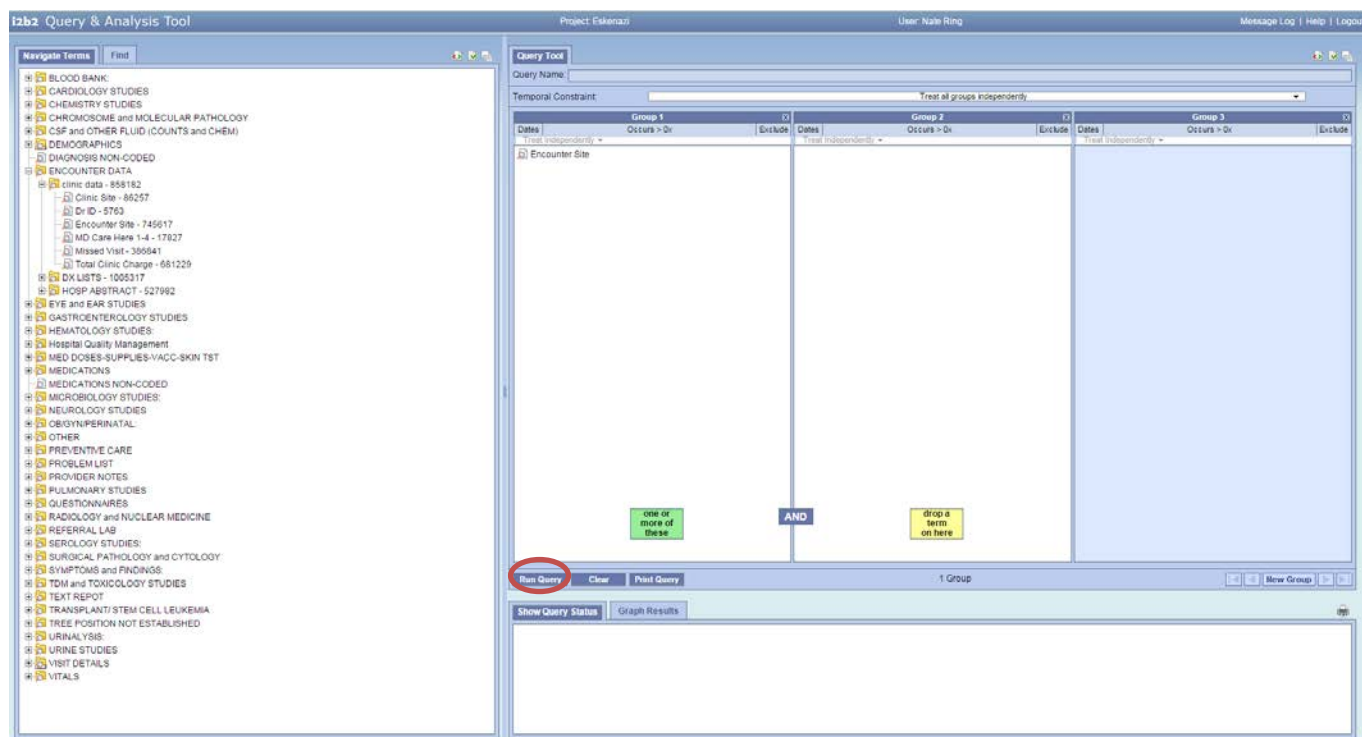
Next you'll drag and drop the icon that says, *Encounter Site* to the Group 1 window. Your screen should look like the images below—in that order.



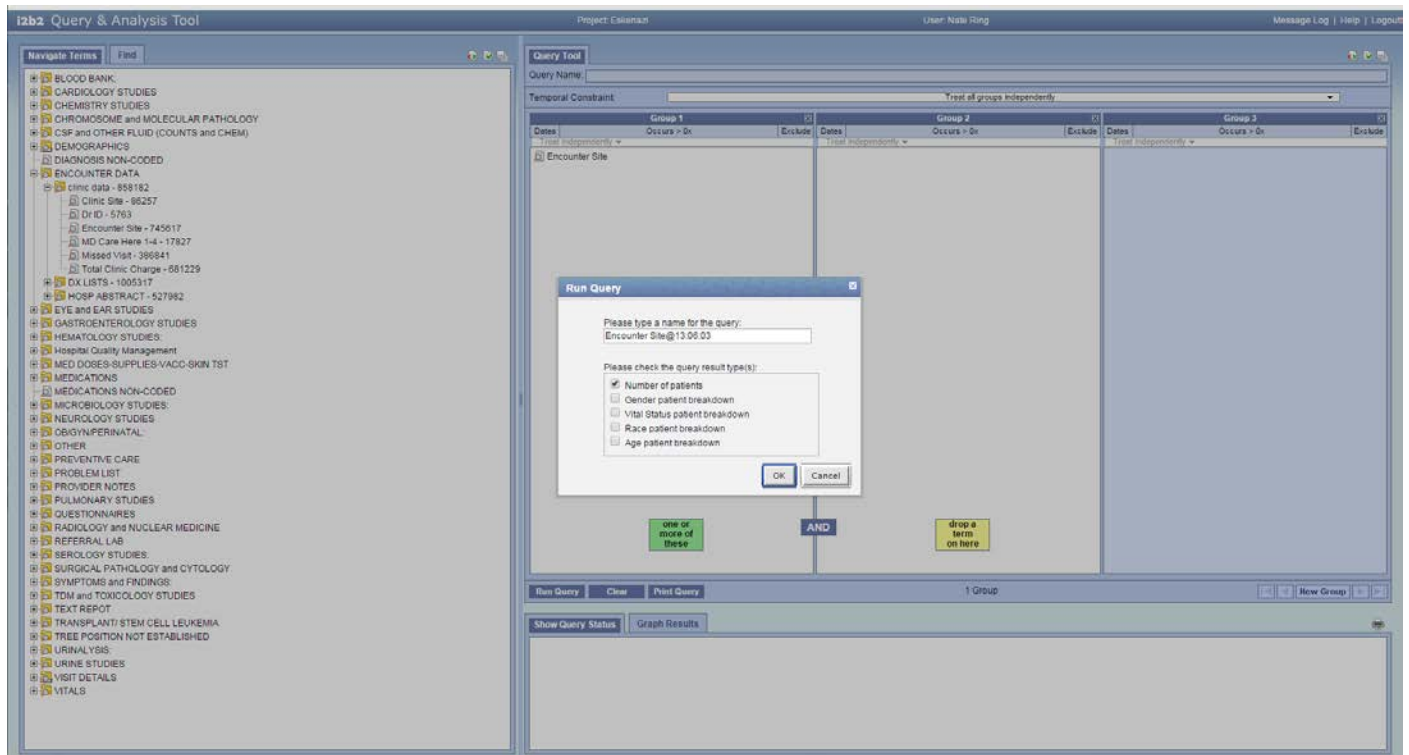
You'll get a pop-up that asks you what search text you want to use. First let's use no value to see the total number of patients with data in that concept. So select *No Value*. Your screen should look like the image below; if it does, then click ok.



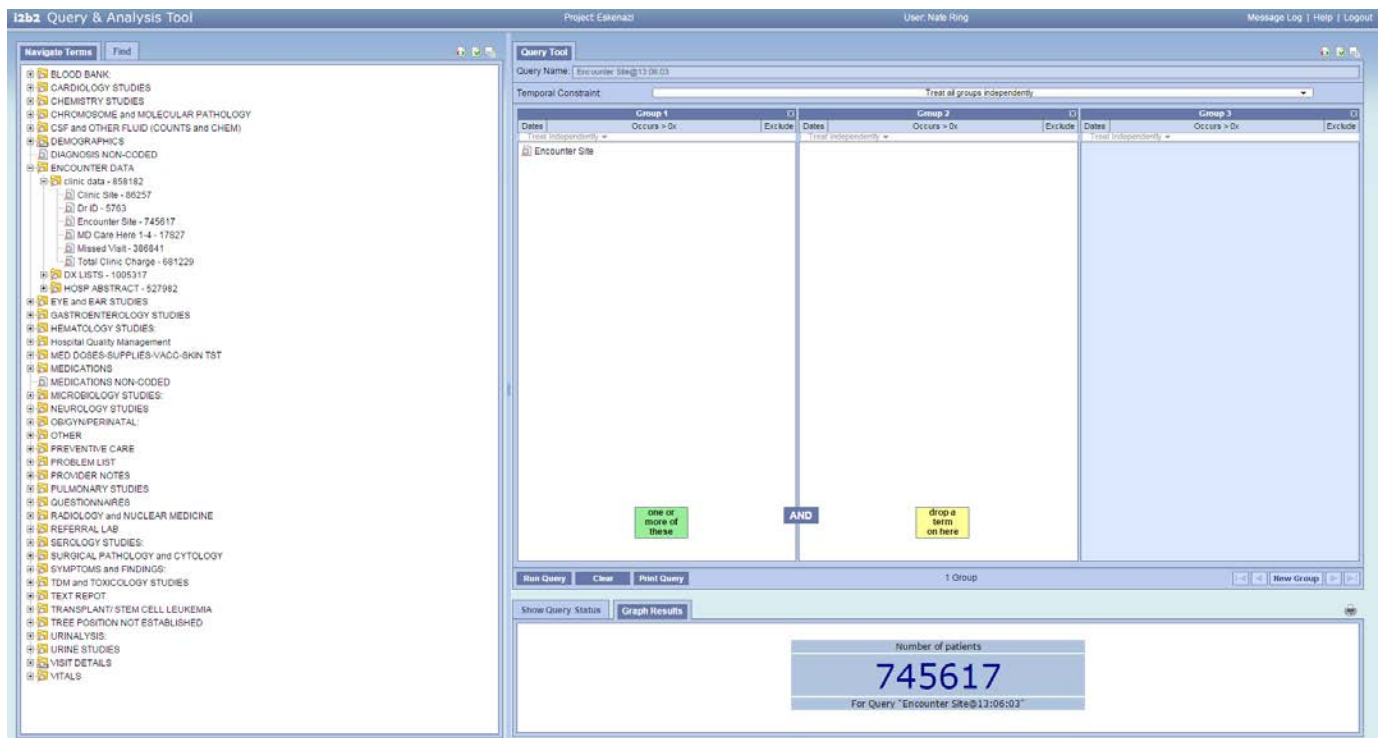
Once you click OK your screen will look like below. Select *Run Query*, which I've circled in red.



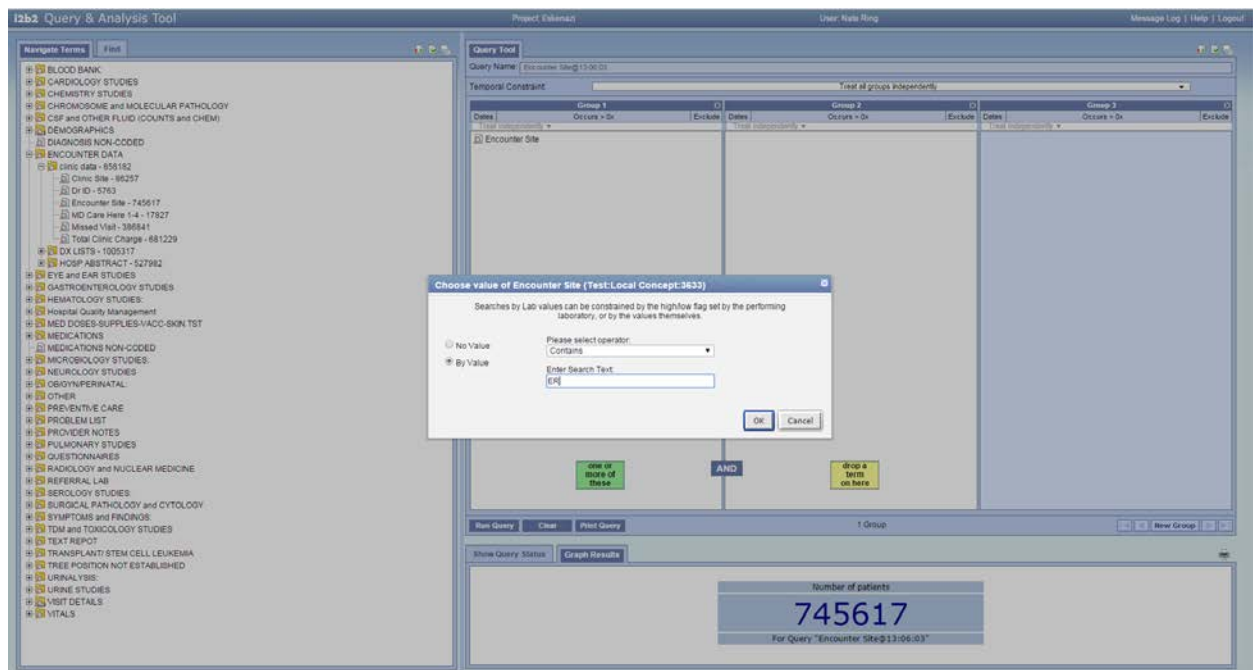
When you click *Run Query* you'll get another pop-up. This will ask you how you'd like your data broken up. For now we are only interested in total number of patients. So just select *OK*. Your screen should look like the image below.



After a few seconds you should get the results at the bottom of the page, a number over 745,000. Your screen should look something like the image below.



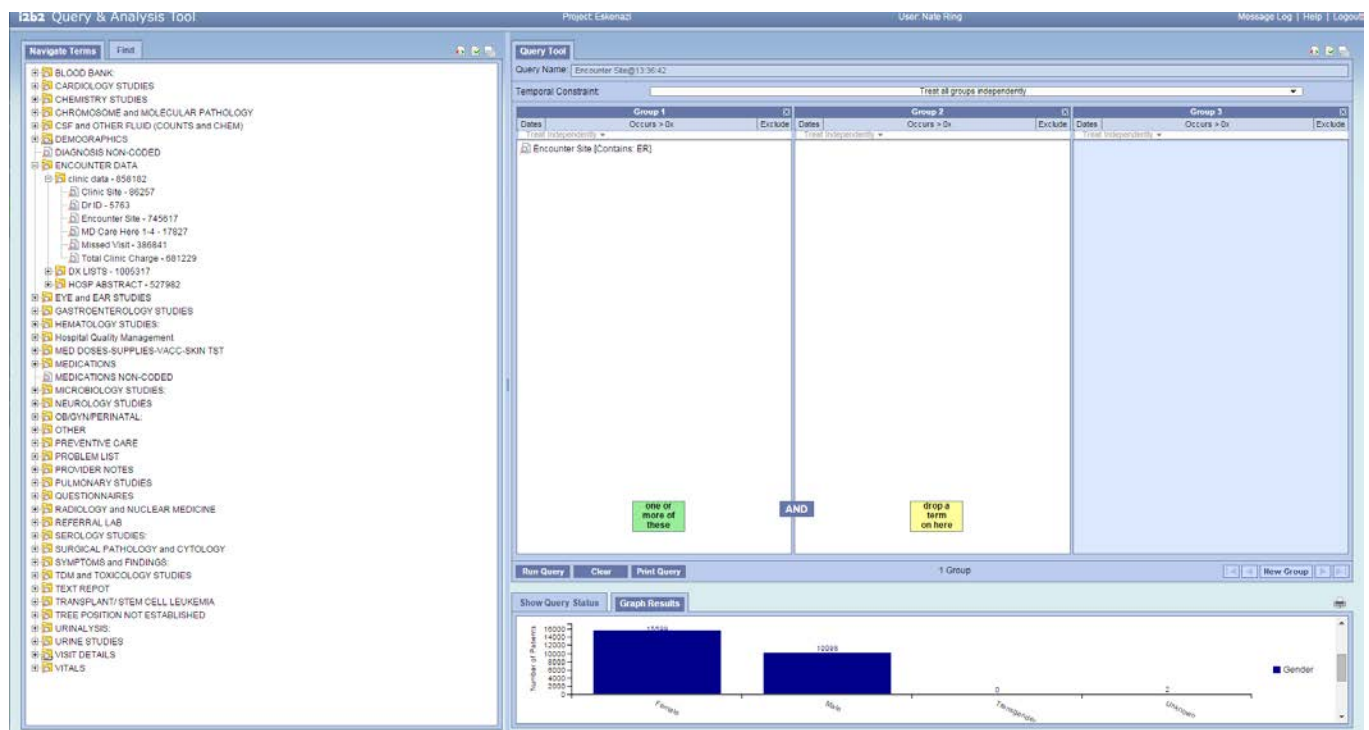
Now we are focused on refining the query to look for ER patients. *Right click* on the concept you dragged over earlier in *Group 1*, then *Set Value...*; you'll see the same window when you dragged the concept over the first time. This time we'll enter "ER" in the field.



You'll notice now the *Group 1* concept has changed to include "Contains: ER". If you mouse over the field you'll see the directory-path for the concept. This can be helpful if you are building more complex

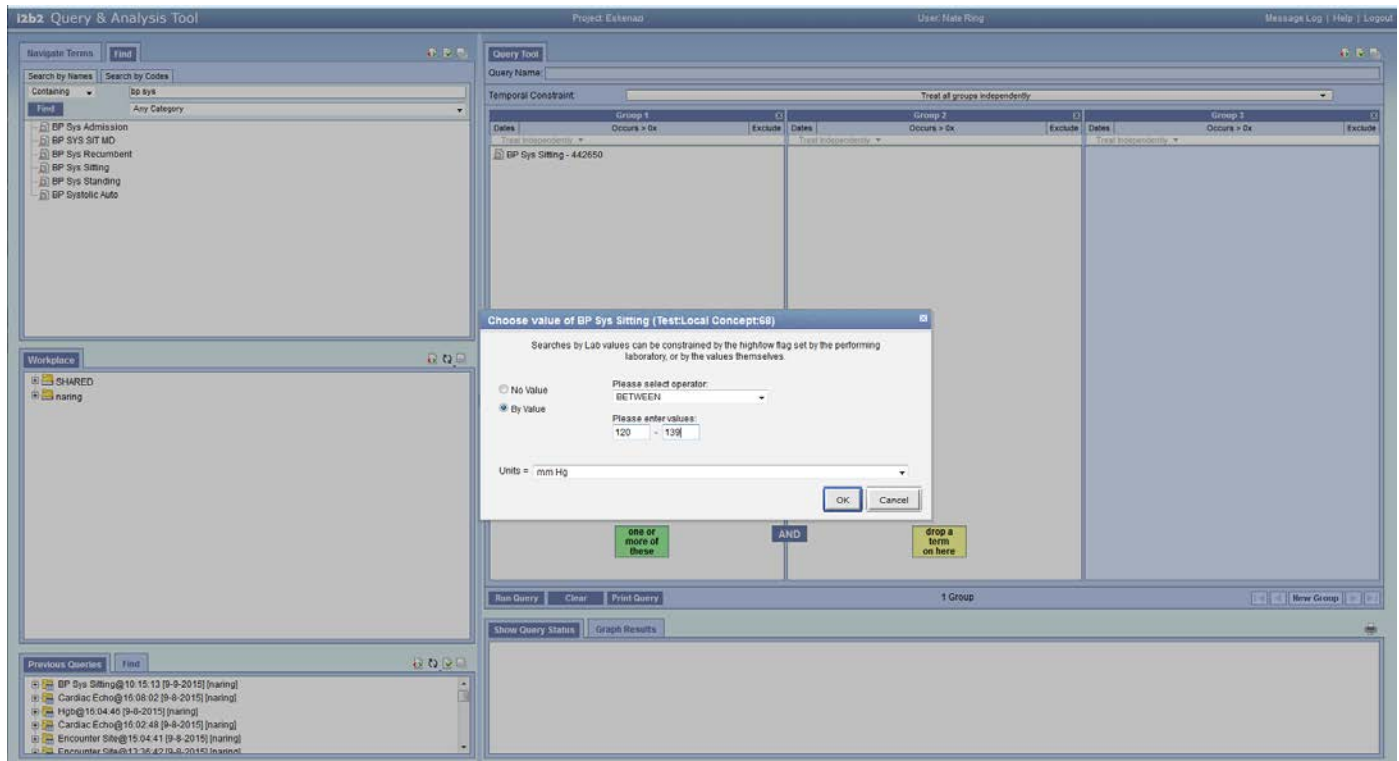
queries. Go ahead and run this query. You'll see your patient count at the bottom—about 25,000 patients. Let's expand on this and now break it down by gender, and age.

Click on *Run Query* again. This time select the *Gender patient breakdown* and *Age patient breakdown*. Your results will be slightly different than before. This time you'll be able choose seeing the data either as bar graphs or tables. The two tabs called *Graph Results* and *Show Query Status* will let you choose how you'd like to display the data.

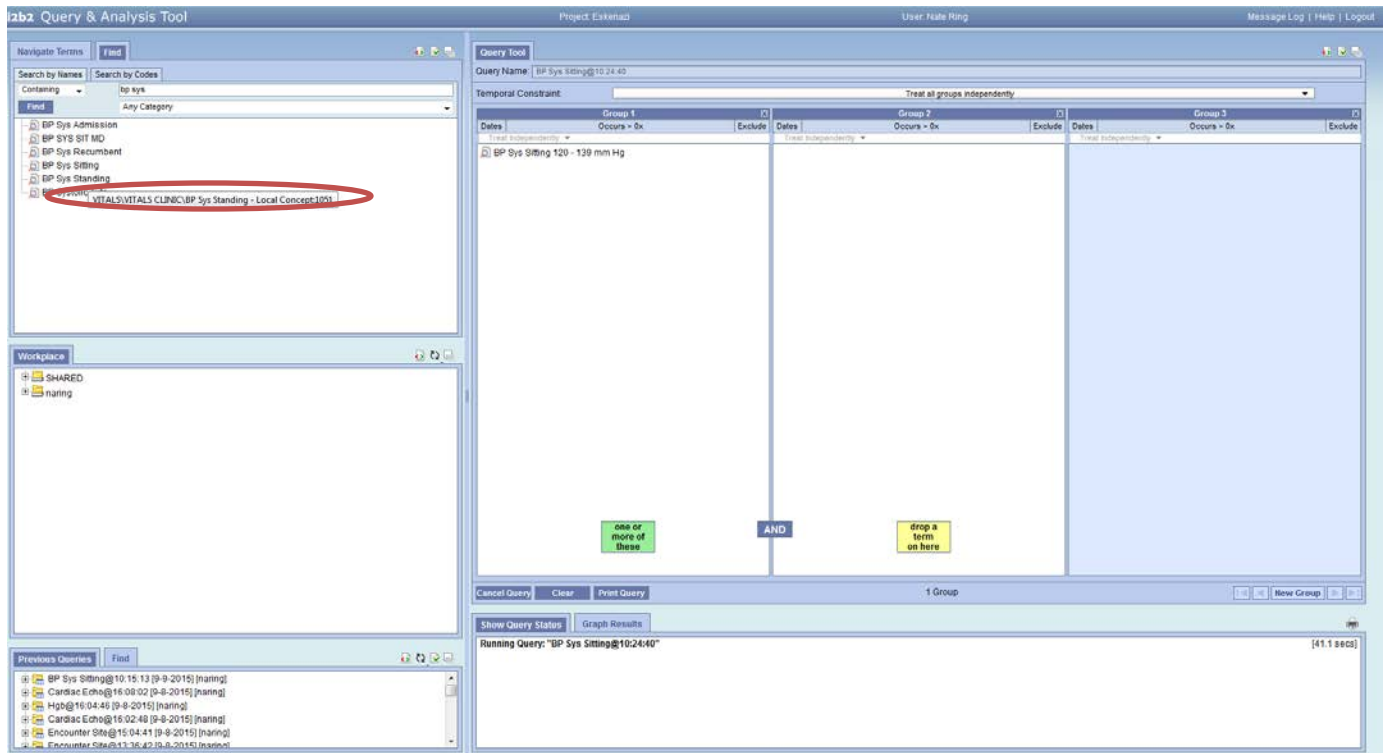


Section two: Labs

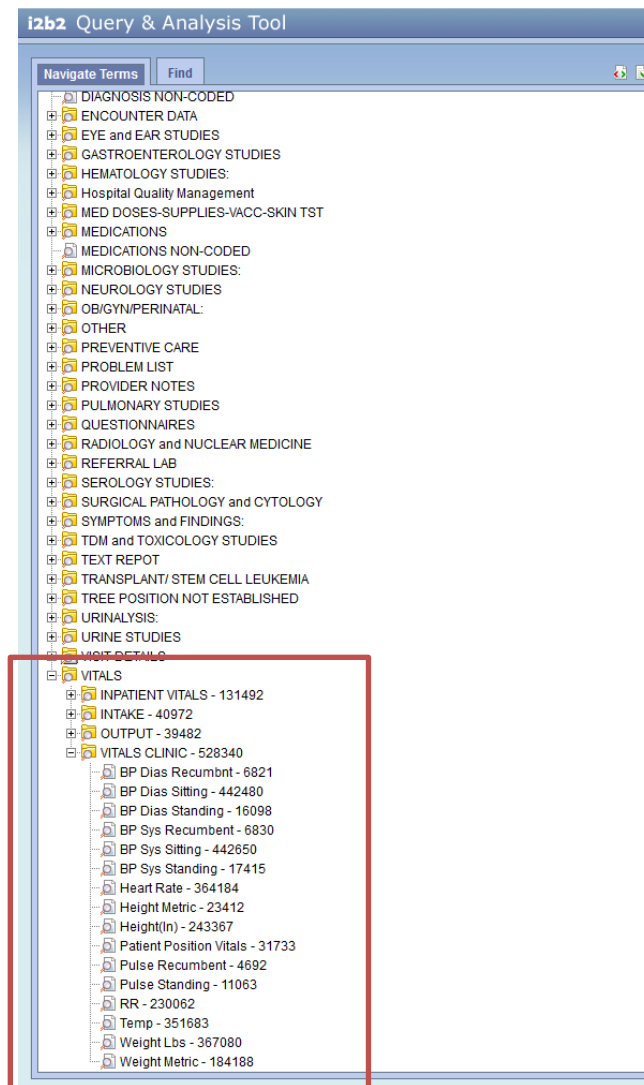
Now that we've played with encounters, let's deal with some lab values, in particular systolic blood pressure while sitting or *BP Sys Sitting*. You can clear the previous work by using the *Clear* button to the right of the *Run Query* button. This time you're not exactly sure where we'd store systolic blood pressure so we'll use the *Find* function—which is located to the top left. We can search for "bp sys" then drag and drop the *BP Sys Sitting* concept to the *Group 1* terms column. Here you'll have a chance to explore the operators: *Greater than*, *Greater than or Equal to*, *Between*, *Less than*, *Less than or equal to*, and *Equal*. You should be familiar with all but the between. The between operates as any integer that is between and including the values you are providing. So if you want to find a normal BP you'll search for 120 and 139. You should get a number around 106,000 or more.



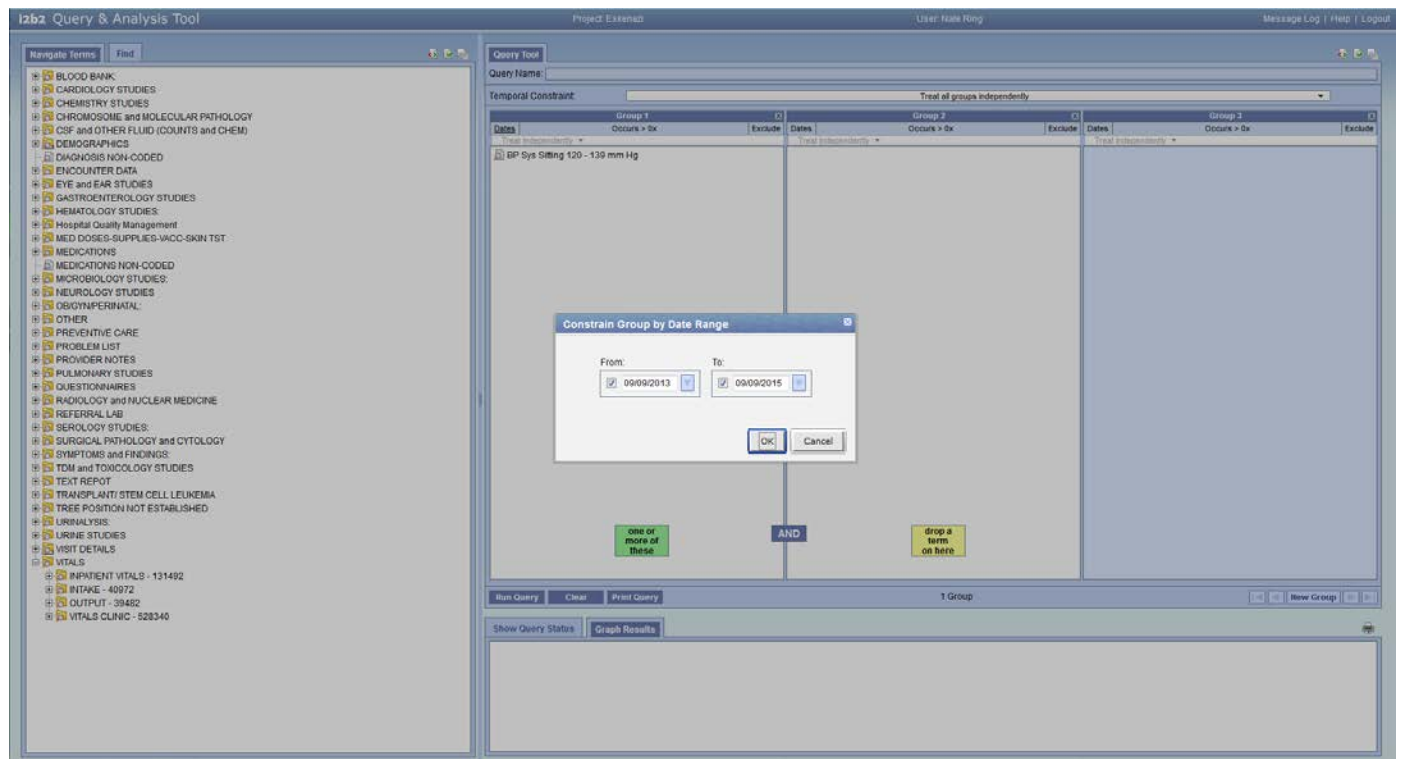
If you were interested in finding similar tests you can mouse over the concept and you'll see its directory path, like the area circled in red below:



VITALS\VITALS CLINIC\BP Sys Sitting. This will tell you how to find or how a particular lab is classed. In this case we see that it's housed in the VITALS CLINIC which also contains several other tests (see image below). We use this to verify we have the right kind of blood pressure test. You'll see that there are many kinds available so take some time to explore the tests.



Assuming we have selected the correct test let's see if we can understand the 106,00 better by seeing how many of those patients have been to Eskenazi in the past 2 years. So select, *Dates* from the *Group 1* box and enter the date range you're interested in making sure you follow MM/DD/YYYY format like I have below. Then click *OK*.



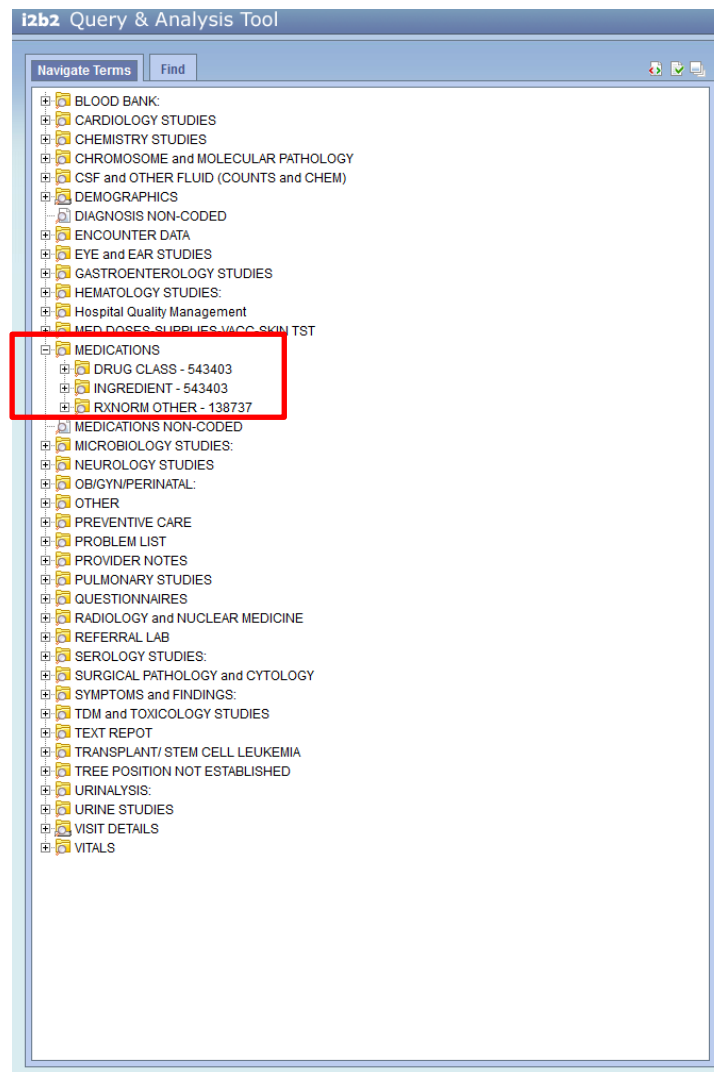
We are also interested in seeing the number of patients who do not have a record of death. So in the *Run Query* select *Vital Status*; then click *Run*. This query may take a few minutes—if it does you'll be prompted and the query will run in the background. To see the results we'll use another feature of i2b2, which is reloading previous queries. The bottom most panel on the left contains all of your previous queries. You can drag and drop the query you built into the *Query Name* section at the very top (circled in red).

The screenshot displays the i2b2 Query & Analysis Tool interface. On the left, the 'Navigate Terms' pane lists various medical categories like BLOOD BANK, CARDIOLOGY STUDIES, and CHEMISTRY STUDIES. Below it, the 'Workplace' pane shows a list of queries, including 'BP Sys Sitting@10:41:19'. The main 'Query Tool' area shows a query for 'BP Sys Sitting@10:41:19' highlighted with a red oval. Below the query, the 'Temporal Constraint' section shows three groups with 'BP Sys Sitting 120 - 139' in the first group. The 'Results' section at the bottom shows a table with 'Number of patients' and the value '8085'.

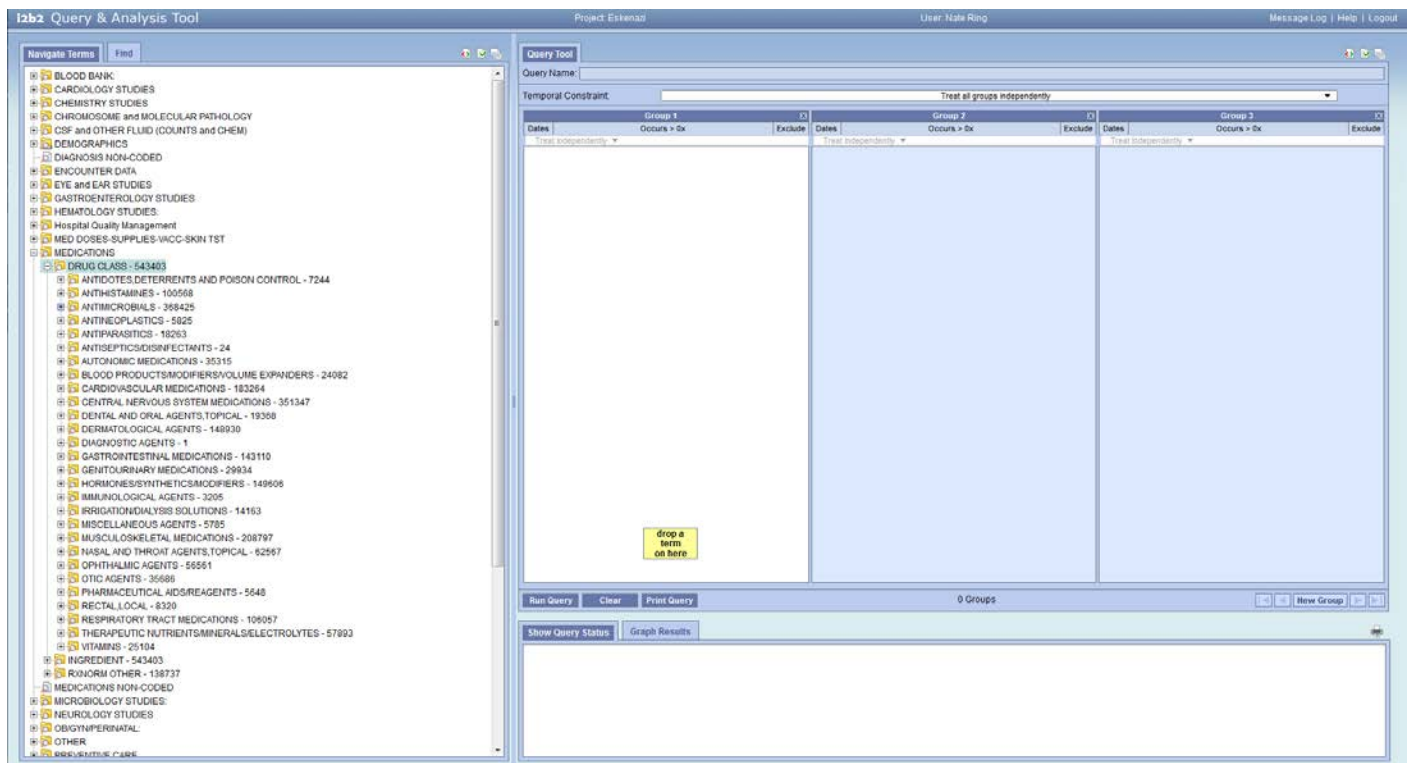
You'll see that you got a number around 8000, and when you scroll through the results you'll see the number of patients with and without a record of death.

Task 3: Medications

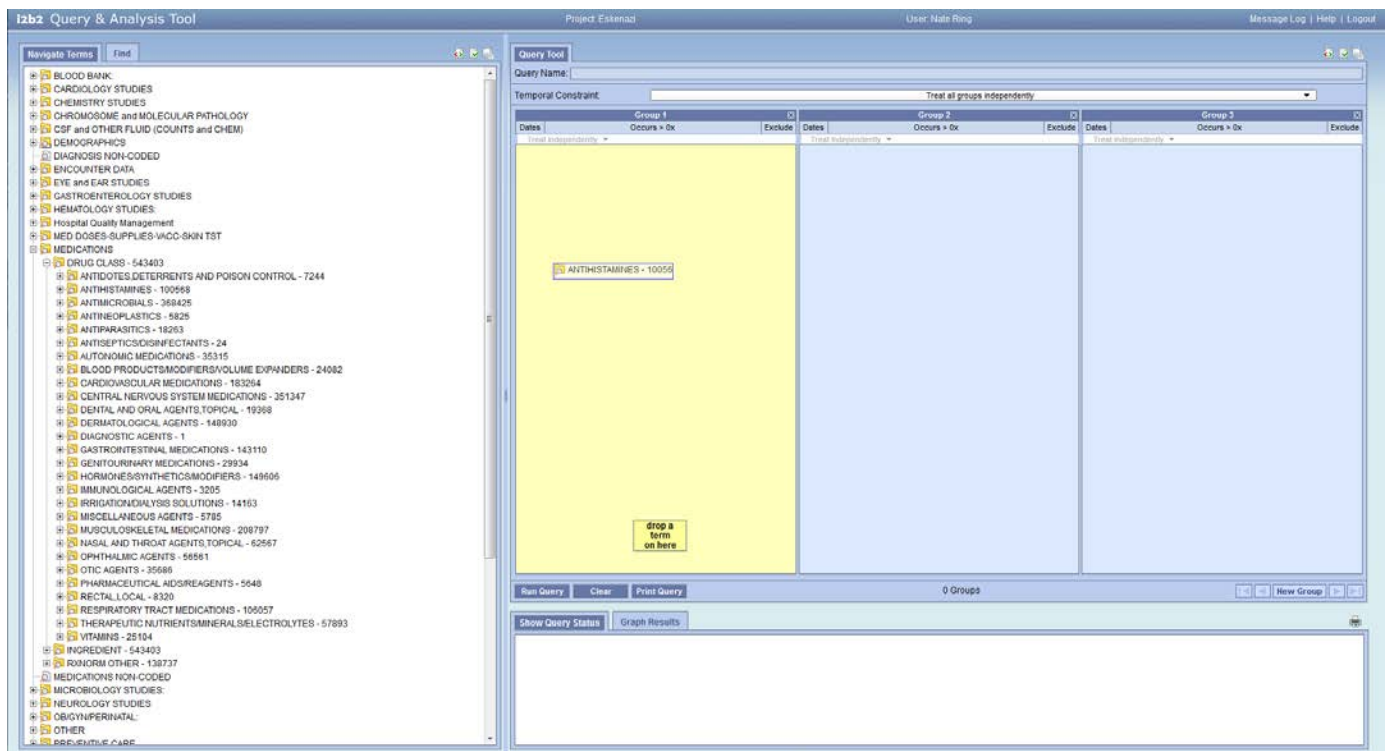
In this section we'll cover the AND and OR functions while finding meds specifically *Acetaminophen* and *Antihistamines*.



It is important to know that i2b2 stores meds in three ways: drug class, ingredient, and RXNORM. Drug class is broken into categories like, “Antihistamines”; the ingredient tab is an alphabetical ordering of generic drug names; finally, RXNORM then provides specific information about drugs including dosage strength. Each system has will have its own merits and pitfalls, so it is a good idea to explore these classing systems. So let’s start with finding *antihistamines*.



Expand the *MEDICATIONS\DRUG CLASS* concept tree. You'll see antihistamines near the top. Since we want the entire group and not any of the sub-classes, we'll just drag and drop the whole *antihistamine* tree.



If we run the query now we should get a number over 100,000 patients. Now because we want a second drug, we can ask many different questions: one OR the other; one AND the other, one AND NOT the other, etc. We'll start with one OR the other then look to see who had both.

l2b2 Query & Analysis Tool

Project: Estenac User: Nate Ring Message Log | Help | Logout

Navigate Terms Find

- PROSPECTIVE STUDIES
- HOSPITAL QUALITY MANAGEMENT
- MED DOSES-SUPPLIES-VACC-SKIN TST
- MEDICATIONS
 - DRUG CLASS - 543403
 - INGREDIENT - 543403
 - ADRUOS - 398466
 - 4-AMINO BENZOATE - 4
 - 6-AMINO CAPROIC ACID - 68
 - ABACAVIR - 123
 - ACAMPROSATE - 54
 - ACARBOSE - 208
 - ACEBUTOLOL - 95
 - ACETAMINOPHEN - 278210
 - ACETAZOLAMIDE - 1261
 - ACETIC ACID - 3066
 - ACETYL CYSTEINE - 656
 - ACITRETIN - 29
 - ACYCLOVIR - 16284
 - ADAPALENE - 277
 - ADEFOVIR - 2
 - ALBUTEROL - 53335
 - ALCLOMETASONE - 139
 - ALENDRONIC ACID - 6851
 - ALFUZOSIN - 139
 - ALLOPURINOL - 7730
 - ALPRAZOLAM - 23009
 - ALPROSTADIL - 7
 - ALTEPLASE - 10
 - ALUMINUM CHLORIDE - 35
 - ALUMINUM HYDROXIDE - 415
 - AMANTADINE - 1781
 - AMCINONIDE - 74
 - AMFOSTINE - 1
 - AMKACIN - 11
 - AMLODIPINE - 355
 - AMINOPIRYLINE - 97
 - AMIODARONE - 2448
 - AMITRIPTYLINE - 30597
 - AMLODIPINE - 41198
 - AMOXAPINE - 6
 - AMOXICILLIN - 169959
 - AMPHETAMINE - 5114
 - AMPHOTERICIN B - 16
 - AMPICILLIN - 5424
 - AMYL NITRITE - 25
 - AMYLASES - 267
 - ANAGRELIDE - 23
 - ANASTROZOLE - 973
 - ANTIPYRINE - 6312
 - APRACLOININE - 8
 - ASCORBIC ACID - 806

Query Tool

Query Name: ANTHISTAMINES@12:27:41

Temporal Constraint: Treat all groups independently

Group 1	Group 2	Group 3
Occurs > 0x	Occurs > 0x	Occurs > 0x
ANTHISTAMINES - 100568		

one or more of these AND drop a term on here

Run Query Clear Print Query 1 Group New Group

Show Query Status Graph Results

Number of patients
100568
For Query "ANTHISTAMINES@12:27:41"

Because *acetaminophen* is an ingredient we'll look there, then ADD it to the same group as *antihistamines*. This will perform an "OR" function.

l2b2 Query & Analysis Tool

Project: Estenac User: Nate Ring Message Log | Help | Logout

Navigate Terms Find

- PROSPECTIVE STUDIES
- HOSPITAL QUALITY MANAGEMENT
- MED DOSES-SUPPLIES-VACC-SKIN TST
- MEDICATIONS
 - DRUG CLASS - 543403
 - INGREDIENT - 543403
 - ADRUOS - 398466
 - 4-AMINO BENZOATE - 4
 - 6-AMINO CAPROIC ACID - 68
 - ABACAVIR - 123
 - ACAMPROSATE - 54
 - ACARBOSE - 208
 - ACEBUTOLOL - 95
 - ACETAMINOPHEN - 278210
 - ACETAZOLAMIDE - 1261
 - ACETIC ACID - 3066
 - ACETYL CYSTEINE - 656
 - ACITRETIN - 29
 - ACYCLOVIR - 16284
 - ADAPALENE - 277
 - ADEFOVIR - 2
 - ALBUTEROL - 53335
 - ALCLOMETASONE - 139
 - ALENDRONIC ACID - 6851
 - ALFUZOSIN - 139
 - ALLOPURINOL - 7730
 - ALPRAZOLAM - 23009
 - ALPROSTADIL - 7
 - ALTEPLASE - 10
 - ALUMINUM CHLORIDE - 35
 - ALUMINUM HYDROXIDE - 415
 - AMANTADINE - 1781
 - AMCINONIDE - 74
 - AMFOSTINE - 1
 - AMKACIN - 11
 - AMLODIPINE - 355
 - AMINOPIRYLINE - 97
 - AMIODARONE - 2448
 - AMITRIPTYLINE - 30597
 - AMLODIPINE - 41198
 - AMOXAPINE - 6
 - AMOXICILLIN - 169959
 - AMPHETAMINE - 5114
 - AMPHOTERICIN B - 16
 - AMPICILLIN - 5424
 - AMYL NITRITE - 25
 - AMYLASES - 267
 - ANAGRELIDE - 23
 - ANASTROZOLE - 973
 - ANTIPYRINE - 6312

Query Tool

Query Name: ANTHISTAMINES@12:27:41

Temporal Constraint: Treat all groups independently

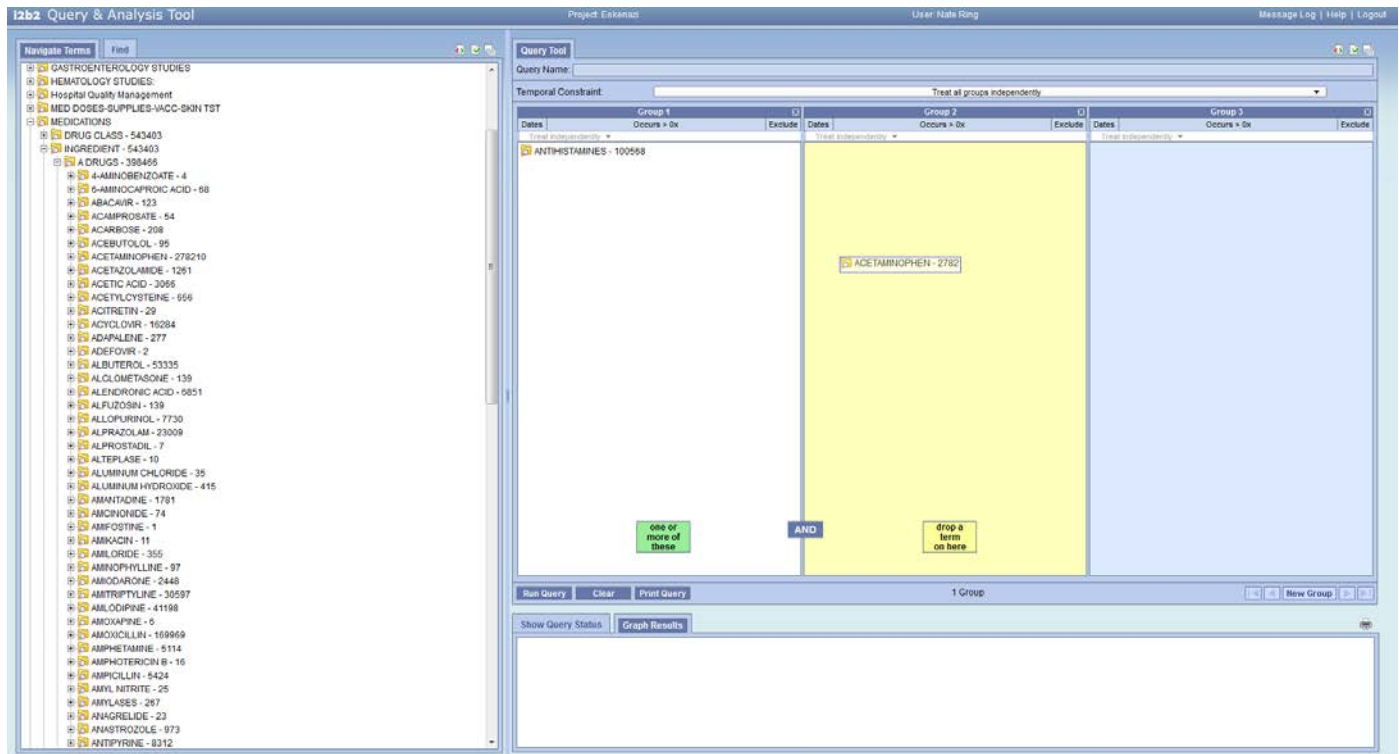
Group 1	Group 2	Group 3
Occurs > 0x	Occurs > 0x	Occurs > 0x
ANTHISTAMINES - 100568	ACETAMINOPHEN - 2782	

one or more of these AND drop a term on here

Run Query Clear Print Query 1 Group New Group

Show Query Status Graph Results

You should get about 300,000 patients. To see who has had both medications we *delete* the *acetaminophen* concept from the first group then drag it to the second group window.

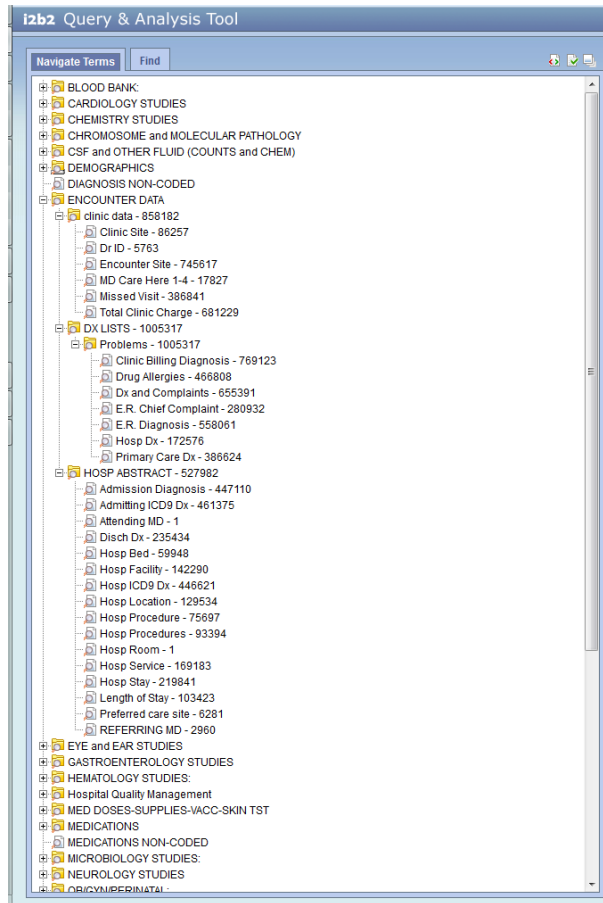


You should get a number around 70,000. You can expand these types of searches to include many more drugs or concepts. We'll expand on this idea in task 4.

Task 4: Diagnoses

By now you should understand most of the interface functions and the drag-and-drop nature of query mechanism. However, many preliminary searches focus on a diagnosis and presence of some other thing. So let's look for an *E.R. Chief Complaint of pain*.

Rather than searching for a concept, *pain* or an icd-9 code of pain we have to look for the type of diagnosis. To illustrate see the image below:



CAVEAT: It is important to point out, some of these concepts will let you enter terms as text, icd-9, or is just a diagnosis type class, you should explore these to find which works best for the project at hand.

Pull over *E.R. Chief Complaint* to the first group. Here we'll enter: *pain*. That is it!

12b2 Query & Analysis Tool Project Eatenazi User: Nate Ring Message Log | Help | Logout

Navigate Terms Find

- BLOOD BANK
- CARDIOLOGY STUDIES
- CHEMISTRY STUDIES
- CHROMOSOME and MOLECULAR PATHOLOGY
- CSF and OTHER FLUID (COUNTS and CHEM)
- DEMOGRAPHICS
- DIAGNOSIS NON-CODED
- ENCOUNTER DATA
 - clinic data - 858182
 - Clinic Site - 85257
 - Dr ID - 5783
 - Encounter Site - 745617
 - MD Care Here 1-4 - 17827
 - Missed Visit - 380841
 - Total Clinic Charge - 681229
 - Problems - 1005317
 - Clinic Billing Diagnosis - 768123
 - Drug Allergies - 466808
 - Dx and Complaints - 655391
 - E.R. Chief Complaint - 280932
 - E.R. Diagnosis - 558061
 - Hosp Dx - 172576
 - Primary Care Dx - 386624
- HOSP ABSTRACT - 527982
- EYE and EAR STUDIES
- GASTROENTEROLOGY STUDIES
- HEMATOLOGY STUDIES
- Hospital Quality Management
- MED DOSES-SUPPLIES-VACC-SKIN TST
- MEDICATIONS
 - MEDICATIONS NON-CODED
- MICROBIOLOGY STUDIES
- NEUROLOGY STUDIES
- OB/GYN/PERINATAL
- OTHER
- PREVENTIVE CARE
- PROBLEM LIST
- PROVIDER NOTES
- PULMONARY STUDIES
- QUESTIONNAIRES
- RADIOLOGY and NUCLEAR MEDICINE
- REFERRAL LAB
- SEROLOGY STUDIES
- SURGICAL PATHOLOGY and CYTOLOGY
- SYMPTOMS and FINDINGS
- TDM and TOXICOLOGY STUDIES
- TEXT REPORT
- TRANSPLANT/ STEM CELL LEUKEMIA
- TREE POSITION NOT ESTABLISHED
- UNKNOWN

Query Tool

Query Name: E.R. Chief Comp@13:51:55

Temporal Constraint: Treat all groups independently

Group 1		Group 2		Group 3	
Dates	Occurs > Dx	Dates	Occurs > Dx	Dates	Occurs > Dx
E.R. Chief Complaint [Contains: pain]					

one or more of these AND drop a term on here

Run Query Clear Print Query 1 Group New Group

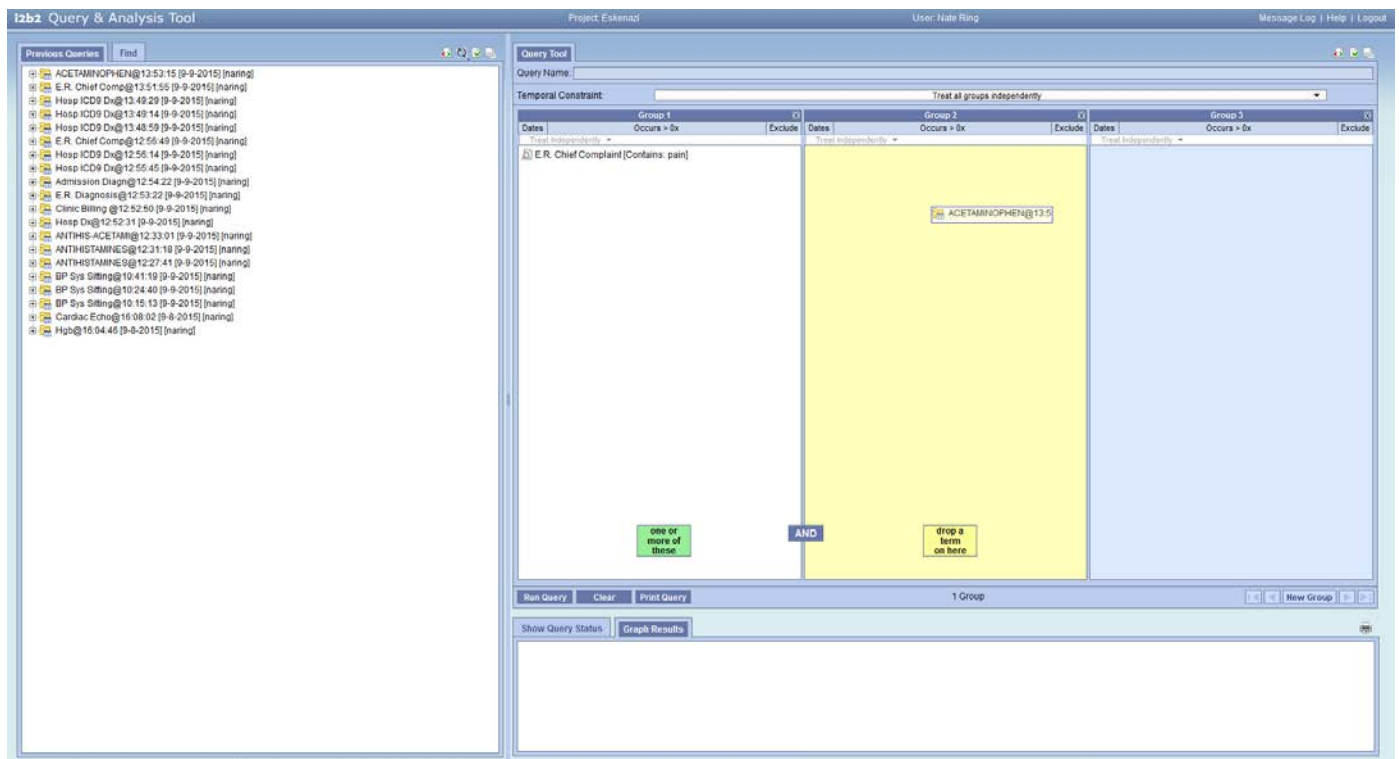
Show Query Status Graph Results

Number of patients

11229

For Query "E.R. Chief Comp@13:51:55"

We can then add in the previous query where we searched for *acetaminophen* from the previous query history.



Now you should be well on your way to building exciting and new queries!

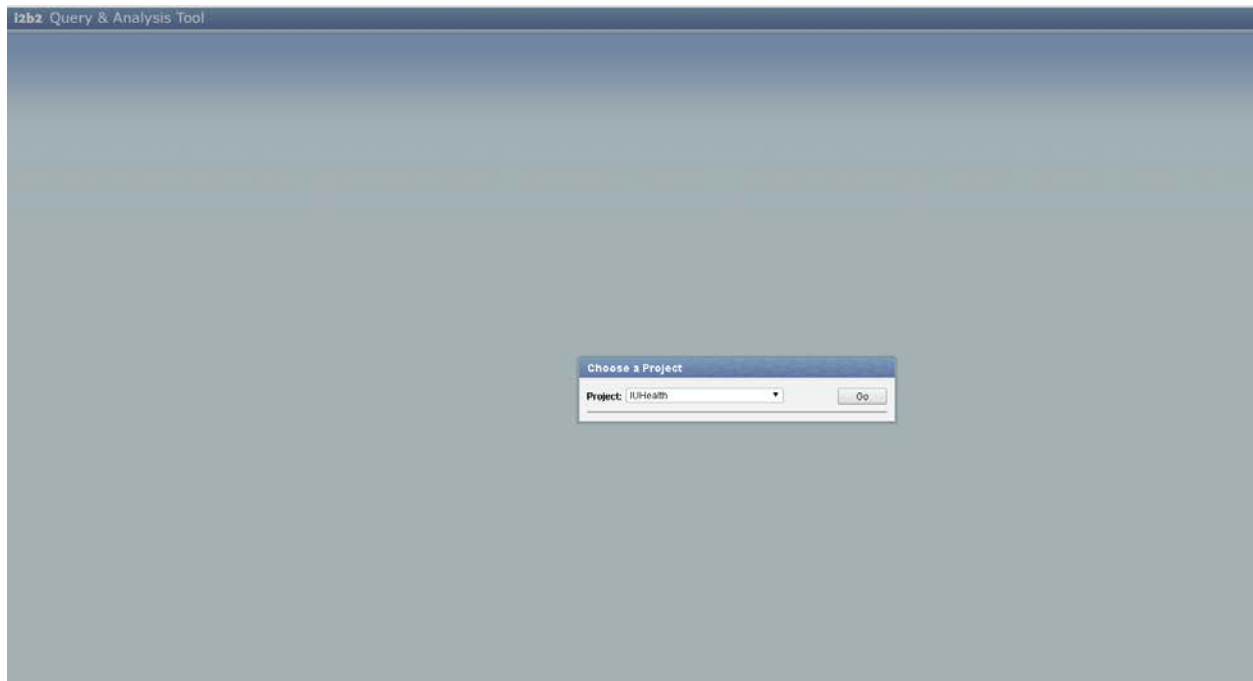
Two final points if you are not interested in moving forward with the IUH section of the materials: First, this data is based on clinical data. Which means that there may be a great deal of variation in how data are stored or where it is located in terms of the concept tree or the type of information you're looking for; second; if data looks very unusual it may not be what you're querying rather something related to how the data is stored. If you find these cases, contact i2b2ctsi@iupui.edu

IU Health Data

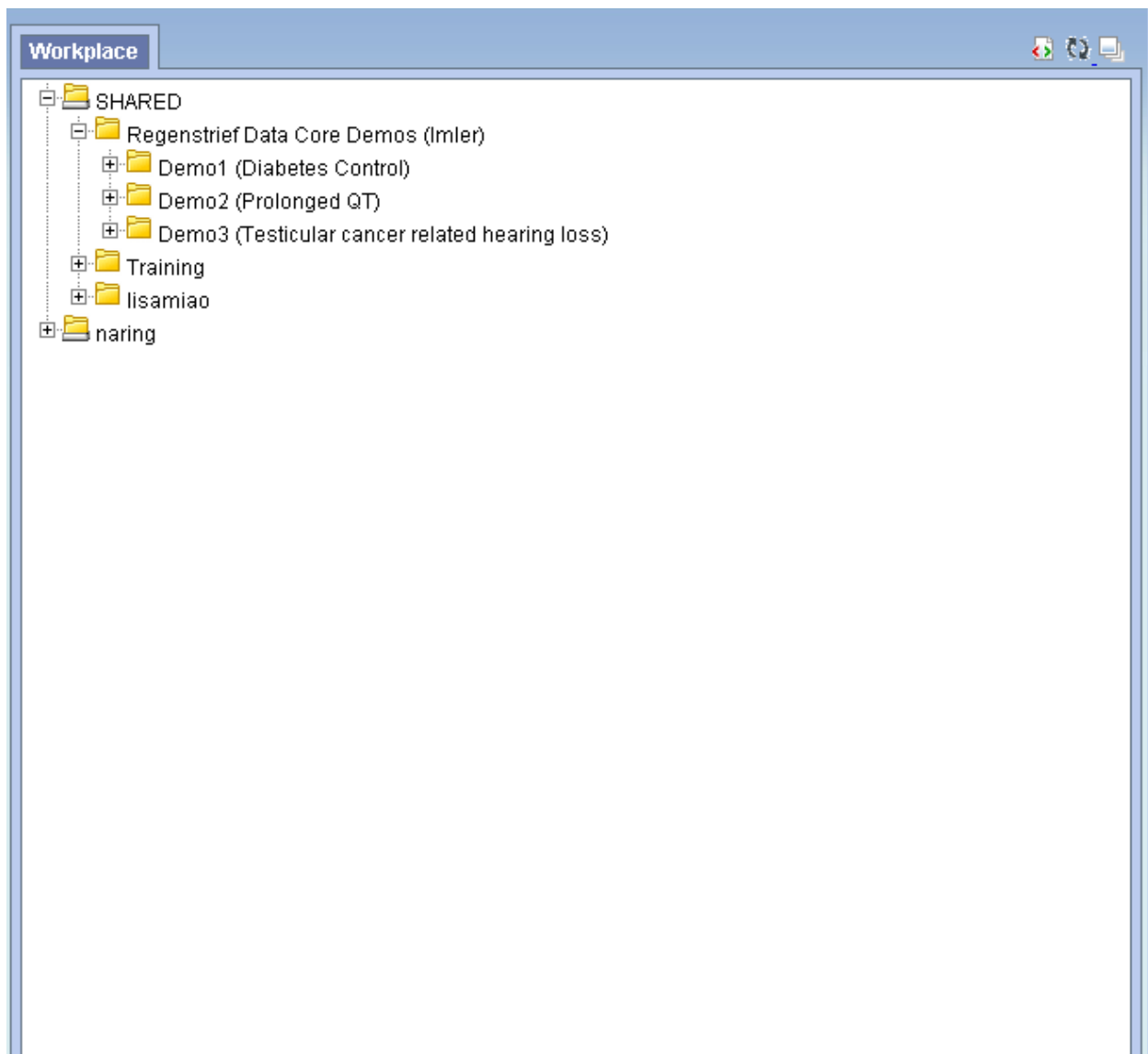
At this point you have decided to explore the data from Indiana University Health. There are several distinctions between IUH and ESK data. First, there are many more patients in this *silo*. This is because IUH is a network of hospitals across the state of Indiana. Second, IUH does not send as much data to the INPC as Eskenazi. So, this silo is somewhat more specialized and selective in its use for feasibility. For example, IUH data is very strong in LABS, CARDIO, report types, and ICD-9's. It is not very good at medications, procedures (unless identified by ICD-9), or *local concepts*. Therefore it is generally a good idea to test in BOTH silos or submit a feasibility request to the Regenstrief Data Core.

Unlike the previous walk-through, since we learned much about the interface and searching we'll focus on expanding queries. For this section, we'll rely on pre-built queries and reading the results.

Here you'll need to log in to the IUH warehouse.



After you've selected IUH, we'll focus on the window of *Shared* folder, which you'll see in the screenshot below.



I've already drilled down on the Regenstrief Data Core demo's for you and the Data Core as built the queries.

Let's explore the first, Diabetes control.

Task 1: Diabetic Retinopathy

We are looking for patients taking metformin and have an uncontrolled diabetic event. So to start we'll look at patients taking Metformin.

Query Tool

Query Name: METFORMIN@08:29:44

Temporal Constraint: Treat all groups independently

Group 1	Group 2	Group 3
<div>Dates Occurs > 0x Exclude</div> <div>Treat Independently</div> <div>METFORMIN</div> <div>one or more of these</div>	<div>Dates Occurs > 0x Exclude</div> <div>Treat Independently</div> <div>drop a term on here</div>	<div>Dates Occurs > 0x Exclude</div> <div>Treat Independently</div>

AND

Run Query Clear Print Query 1 Group New Group

Show Query Status Graph Results

Number of patients

93036

For Query "METFORMIN@08:29:44"

If you click on the printer icon on the bottom right, you'll see the demographic break down of patients taking Metformin (see below).

Query Name: No Query Name is currently provided. Temporal Constraint: Treat all groups independently		
Group 1		
Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
Medication\Ingredient\MI Drugs\METFORMIN	METFORMIN	

Finished Query: "METFORMIN@08:29:44"

[22 secs]

Number of patients for "METFORMIN@08:29:44"

patient_count: 93036

Gender patient breakdown for "METFORMIN@08:29:44"

Female: 50671

Male: 42176

Transgender: 0

Unknown: 5

Race patient breakdown for "METFORMIN@08:29:44"

American Indian or Alaska Native: 65

Asian: 837

Black or African American: 12568

Hispanic or Latino: 251

Unknown or Other: 1772

White: 76104

Age patient breakdown for "METFORMIN@08:29:44"

0-9 years old: 6

10-17 years old: 208

18-34 years old: 3984

35-44 years old: 6934

45-54 years old: 13500

55-64 years old: 23078

65-74 years old: 26334

75-84 years old: 14493

>= 65 years old: 45142

>= 85 years old: 4315

Not recorded: 0

Here there are a few things to pay attention to: cohort design, concept paths, patient counts, and demographics. First, you'll see at the very top, we've specified no date restrictions, no exclusion criteria, no required number of occurrences, and no temporal constraints. Put in other words, this grabs all patients with a record of having Metformin as found in the MEDICATIONS section of i2b2. If you were to add additional *METFORMIN* concepts, they too would be listed here and likely the numbers would increase across the board.

Now we want to find the number of patients with an A1C over 7%. So drag over the "Number of patients on metformin with HgbA1c >=7" folder to the query name section of i2b2. Your results should look like the image below.

Query Name: No Query Name is currently provided. Temporal Constraint: Treat all groups independently		
Group 1		
Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
Medication\Ingredient\IM Drugs\METFORMIN	METFORMIN	

Finished Query: "METFORMIN@08:29:44"

[22 secs]

Number of patients for "METFORMIN@08:29:44"

patient_count: 93036

Gender patient breakdown for "METFORMIN@08:29:44"

Female: 50671

Male: 42176

Transgender: 0

Unknown: 5

Race patient breakdown for "METFORMIN@08:29:44"

American Indian or Alaska Native: 65

Asian: 837

Black or African American: 12568

Hispanic or Latino: 251

Unknown or Other: 1772

White: 76104

Age patient breakdown for "METFORMIN@08:29:44"

0-9 years old: 6

10-17 years old: 208

18-34 years old: 3984

35-44 years old: 6934

45-54 years old: 13500

55-64 years old: 23078

65-74 years old: 26334

75-84 years old: 14493

>= 65 years old: 45142

>= 85 years old: 4315

Not recorded: 0

Query Name: No Query Name is currently provided. Temporal Constraint: Treat all groups independently		
Group 1		
Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
Medication\Ingredient\IM Drugs\METFORMIN	METFORMIN	
Group 2		
Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
CHEMISTRY STUDIES\CARBOHYDRATE\DIABETES STUDIES\GLYCOSYLATED HGB TESTS\Hgb A1C Bld Qn HPLC	Hgb A1C Bld Qn HPLC >= 7 % Hgb	GE : 7 % Hgb

Finished Query: "Number of patients on metformin with HgbA1c >=7"

[73 secs]

Ctnd Below:

Finished Query: "Number of patients on metformin with HgbA1c >=7"

[73 secs]

Number of patients for "Number of patients on metformin with HgbA1c >=7"

patient_count: 22338

Gender patient breakdown for "Number of patients on metformin with HgbA1c >=7"

Female: 11022

Male: 11253

Transgender: 0

Unknown: 0

Race patient breakdown for "Number of patients on metformin with HgbA1c >=7"

American Indian or Alaska Native: 20

Asian: 243

Black or African American: 4131

Hispanic or Latino: 31

Unknown or Other: 269

White: 17564

Age patient breakdown for "Number of patients on metformin with HgbA1c >=7"

0-9 years old: 0

10-17 years old: 6

18-34 years old: 310

35-44 years old: 1221

45-54 years old: 3303

55-64 years old: 6056

65-74 years old: 6557

75-84 years old: 3664

>= 65 years old: 11379

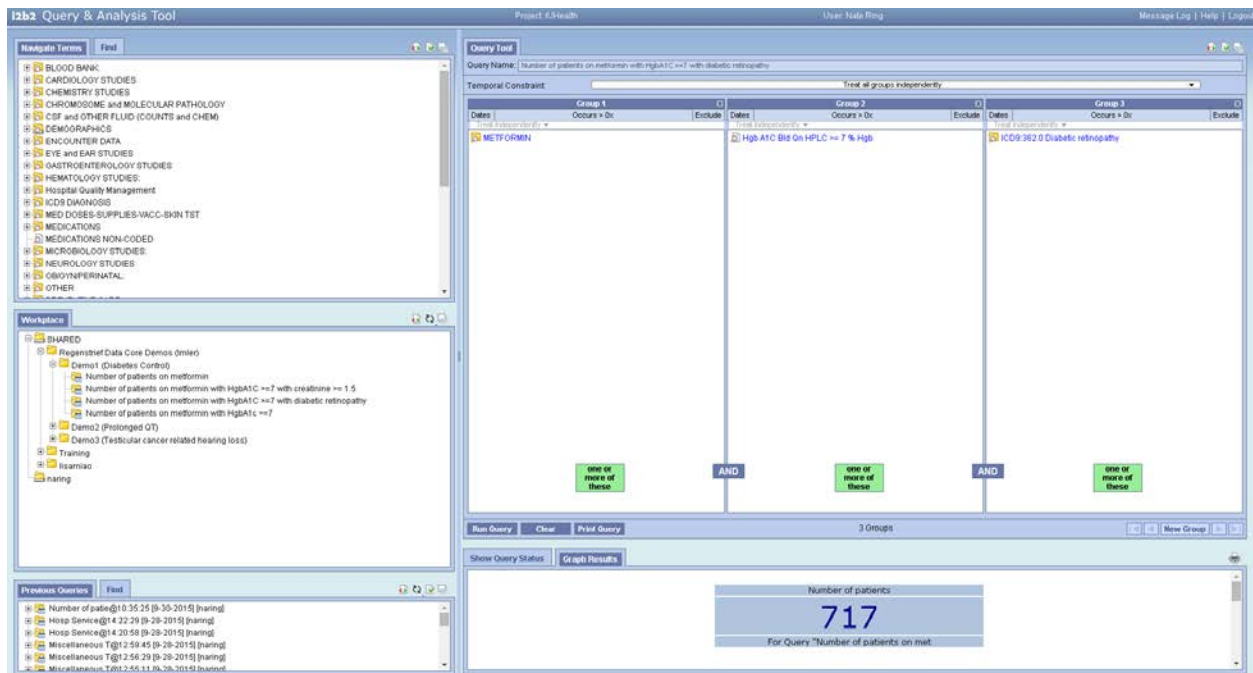
>= 85 years old: 1158

Not recorded: 0

Here you'll see the progressive break-down of the cohort: First patients taking Metformin (Group 1), then of those patients who has had an HgbA1C >=7. This can be a fairly powerful function to see determine where potential relationships between concepts exist. If we were to reverse this cohort, placing hgb A1C in the first group then Metformin in the second box we'd see the number of patients with an A1C >=7, then who has taken Metformin. This output function is consistent as the number of group's increases so it may be worth exploring.

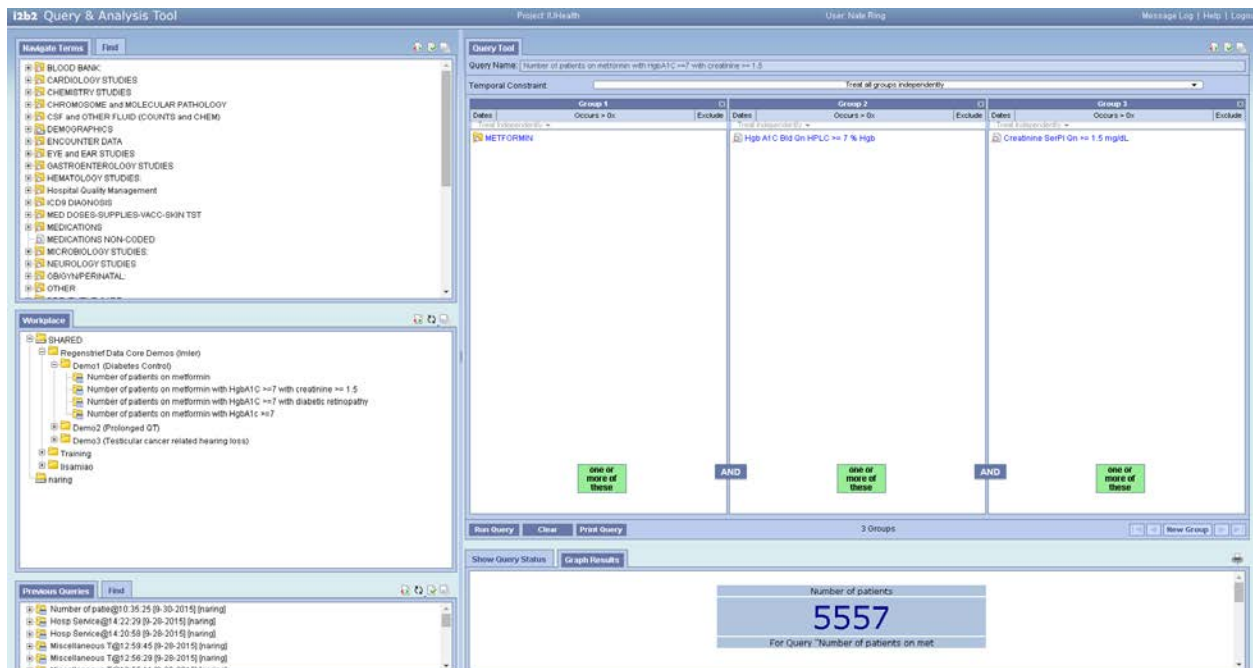
We now want to explore patients who may have diabetic retinopathy; we can do this in two ways. First we can look for patients who have the icd-9 code, then we can look for those who have creatinine levels greater than 1.5 mg/dL; however we can learn a great deal more if we *Combine* those queries.

First, let's evaluate the diabetic retinopathy query. After dragging the concept over to the *Query Name* section as we've done in the past, your screen should look like the image below.



Here you'll notice that we've just extended the previous query and added the icd-9 for *Diabetic Retinopathy* and your count should be 717 patients. It is important to remember you can evaluate the step-wise breakdown of this query by changing the order of group selection.

Now let's evaluate the second method, diabetic retinopathy via creatinine levels. After dragging the concept over to the *Query Name* section as we've done in the past, your screen should look like the image below.

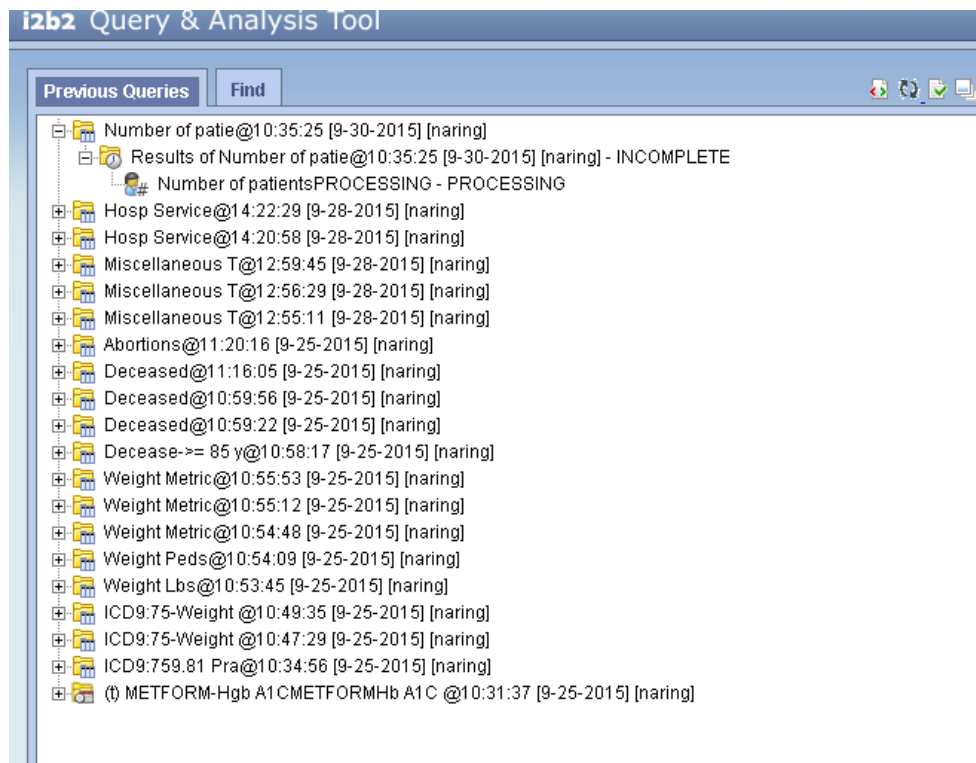


Here we see there are 5557 patients that meet our criteria, but now we want to combine the results and see who meets one or the other condition. Since we have already cued up the results we can show that we find an additional 357 patients by combining the queries, as we see below.

The screenshot displays the i2b2 Query & Analysis Tool interface. The main window is divided into several panes:

- Metadata Items:** A list of medical categories such as BLOOD BANK, CARDIOLOGY STUDIES, CHEMISTRY STUDIES, etc.
- Workplace:** A tree structure showing a hierarchy of queries, including 'Regenstrief Data Core Demos (narr)' and 'Demo1 (Diabetes Control)'.
- Previous Queries:** A list of previously executed queries, including 'Number of patients@10:35:25 [narr]' and 'Results of Number of patients@10:35:25 [narr]'.
- Query Tool:** The central pane for constructing queries. It shows a query named 'Number of patients@10:35:25' with a temporal constraint. The query is built using a combination of 'AND' and 'OR' operators. The 'AND' operator is highlighted in green, and the 'OR' operator is highlighted in yellow.
- Show Query Status:** A pane at the bottom right showing the result of the query. It displays 'Number of patients' as '5914' for the query 'Number of patients@10:35:25'.

It is important to point out, some queries you can build can take quite a while to run. While we address performance issues, it may be helpful to notice when queries have finished. You can do this by expanding your previous queries and expanding the specific query you ran. For example see the image below:



Here you'll notice it tells you that it's still processing and that the query is *incomplete*. You can hit the refresh button I've circled in RED to update this window. At this time there is no progress bar or alert system to let you know when complex queries have finished.

Query Name: No Query Name is currently provided. Temporal Constraint: Treat all groups independently		
Group 1		
Date From: none Date To: none Excluded?: false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
Medication\Ingredient\IM Drugs\METFORMIN	METFORMIN	

Finished Query: "METFORMIN@08:29:44"

[22 secs]

Number of patients for "METFORMIN@08:29:44"

patient_count: 93036

Gender patient breakdown for "METFORMIN@08:29:44"

Female: 50671

Male: 42176

Transgender: 0

Unknown: 5

Race patient breakdown for "METFORMIN@08:29:44"

American Indian or Alaska Native: 65

Asian: 837

Black or African American: 12568

Hispanic or Latino: 251

Unknown or Other: 1772

White: 76104

Age patient breakdown for "METFORMIN@08:29:44"

0-9 years old: 6

10-17 years old: 208

18-34 years old: 3984

35-44 years old: 6934

45-54 years old: 13500

55-64 years old: 23078

65-74 years old: 26334

75-84 years old: 14493

>= 65 years old: 45142

>= 85 years old: 4315

Not recorded: 0

Query Name: No Query Name is currently provided. Temporal Constraint: Treat all groups independently		
Group 1		
Date From: none Date To: none Excluded?: false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
Medication\Ingredient\IM Drugs\METFORMIN	METFORMIN	
Group 2		
Date From: none Date To: none Excluded?: false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
CHEMISTRY STUDIES\CARBOHYDRATE\DIABETES STUDIES\GLYCOSYLATED HGB TESTS\Hgb A1c Bld Qn HPLC	Hgb A1c Bld Qn HPLC >= 7 % Hgb	GE : 7 % Hgb

Finished Query: "Number of patients on metformin with HgbA1c >=7"

[73 secs]

Continued Below

Number of patients for "Number of patients on metformin with HgbA1c >=7"

patient_count: 22338

Gender patient breakdown for "Number of patients on metformin with HgbA1c >=7"

Female: 11022

Male: 11253

Transgender: 0

Unknown: 0

Race patient breakdown for "Number of patients on metformin with HgbA1c >=7"

American Indian or Alaska Native: 20

Asian: 243

Black or African American: 4131

Hispanic or Latino: 31

Unknown or Other: 269

White: 17564

Age patient breakdown for "Number of patients on metformin with HgbA1c >=7"

0-9 years old: 0

10-17 years old: 6

18-34 years old: 310

35-44 years old: 1221

45-54 years old: 3303

55-64 years old: 6056

65-74 years old: 6557

75-84 years old: 3664

>= 65 years old: 11379

>= 85 years old: 1158

Not recorded: 0

Query Name: No Query Name is currently provided.
Temporal Constraint: Treat all groups independently

Group 1

Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently

Path	Concept/Term	Other Information
undefined	undefined	
undefined	undefined	

Finished Query: "Number of patie@10:35:25"

[1482 secs]

Number of patients for "Number of patie@10:35:25"

patient_count: 5914

Here I want to point your attention to two important things. First, the cohort design; second, the time required to run.

First, you'll see that it combines the *Structure* of each query. This means that since we kept *Metformin* and A1C constant, but varied the third group (creatinine OR icd-9), i2b2 runs the query as those queries as we've previously done however adds the OR status for the creatinine and icd-9 conditions. Because we chose to combine our query in this manner, we must shift our attention to the resultant time: 1482 seconds. Now, you may assume that it takes less time to just make the addition by adding a second OR to the final group, however it in fact can slow down results because i2b2 will need to search for new concepts. For example see the output below.

Query Name: No Query Name is currently provided.
Temporal Constraint: Treat all groups independently

Group 1

Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently

Path	Concept/Term	Other Information
Medication\Ingredient\IM Drugs\METFORMIN	METFORMIN	

Finished Query: "METFORMIN@08:29:44"

[22 secs]

Number of patients for "METFORMIN@08:29:44"

patient_count: 93036

Gender patient breakdown for "METFORMIN@08:29:44"

Female: 50671

Male: 42176

Transgender: 0

Unknown: 5

Race patient breakdown for "METFORMIN@08:29:44"

American Indian or Alaska Native: 65

Asian: 837

Black or African American: 12568

Hispanic or Latino: 251

Unknown or Other: 1772

White: 76104

Age patient breakdown for "METFORMIN@08:29:44"

0-9 years old: 6

10-17 years old: 208

18-34 years old: 3984

35-44 years old: 6934

45-54 years old: 13500

55-64 years old: 23078

65-74 years old: 26334

75-84 years old: 14493

>= 65 years old: 45142

>= 85 years old: 4315

Not recorded: 0

Query Name: No Query Name is currently provided.
Temporal Constraint: Treat all groups independently

Group 1

Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently

Path	Concept/Term	Other Information
Medication\Ingredient\IM Drugs\METFORMIN	METFORMIN	

Group 2

Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently

Path	Concept/Term	Other Information
CHEMISTRY STUDIES\CARBOHYDRATE/DIABETES STUDIES\GLYCOSYLATED HGB TESTS\Hgb A1C Bld Qn HPLC	Hgb A1C Bld Qn HPLC >= 7 % Hgb	GE : 7 % Hgb

Finished Query: "Number of patients on metformin with HgbA1c >=7"

[73 secs]

Finished Query: "Number of patients on metformin with HgbA1c >=7"

[73 secs]

Number of patients for "Number of patients on metformin with HgbA1c >=7"

patient_count: 22338

Gender patient breakdown for "Number of patients on metformin with HgbA1c >=7"

Female: 11022

Male: 11253

Transgender: 0

Unknown: 0

Race patient breakdown for "Number of patients on metformin with HgbA1c >=7"

American Indian or Alaska Native: 20

Asian: 243

Black or African American: 4131

Hispanic or Latino: 31

Unknown or Other: 269

White: 17564

Age patient breakdown for "Number of patients on metformin with HgbA1c >=7"

0-9 years old: 0

10-17 years old: 6

18-34 years old: 310

35-44 years old: 1221

45-54 years old: 3303

55-64 years old: 6056

65-74 years old: 6557

75-84 years old: 3664

>= 65 years old: 11379

>= 85 years old: 1158

Not recorded: 0

Query Name: No Query Name is currently provided. Temporal Constraint: Treat all groups independently		
Group 1		
Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
undefined	undefined	
undefined	undefined	

Finished Query: "Number of patie@10:35:25"

[1482 secs]

Number of patients for "Number of patie@10:35:25"

patient count: 5914

Query Name: No Query Name is currently provided. Temporal Constraint: Treat all groups independently		
Group 1		
Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
Medication\Ingredient\Drugs\METFORMIN	METFORMIN	
Group 2		
Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
CHEMISTRY STUDIES\CARBOHYDRATE/DIABETES STUDIES\GLYCOSYLATED HGB TESTS\Hgb A1C Bld Qn HPLC	Hgb A1C Bld Qn HPLC >= 7 % Hgb	GE : 7 % Hgb
Group 3		
Date From: none Date To: none Excluded? false Occurs X times: > 0 Relevance %: 100 Temporal Constraint: Treat Independently		
Path	Concept/Term	Other Information
CHEMISTRY STUDIES\ROUTINE SER-PLAS CHEM TESTS\CREATININE TESTS\Creatinine SerPI Qn	Creatinine SerPI Qn >= 1.5 mg/dL	GE : 1.5 mg/dL
Diseases of the nervous system and sense organs \ Disorders of the eye and adnexa \ Other retinal disorders \ Diabetic retinopathy	ICD9:362.0 Diabetic retinopathy	

Finished Query: "METFO-Hgb A-Creat@11:20:19"

[1009 secs]

Number of patients for "METFO-Hgb A-Creat@11:20:19"
patient_count: 5914

Here you see that the overall time was increased by 1009 seconds. Which brings up a caveat about the i2b2 program. Some times it's easier to build on previously run queries as the results are cached rather than extending the existing query.

Now that we have learned about some of the more advanced features, lets explore a different topic.

Task 2: Long QT

Here we are interested in individuals who have taken a medication and then develop a long QT. We will explore the event-ordered function of i2b2.

Pull over the previously run query: “Number of patients with QTC \geq 480”, your screen should look like the following image.

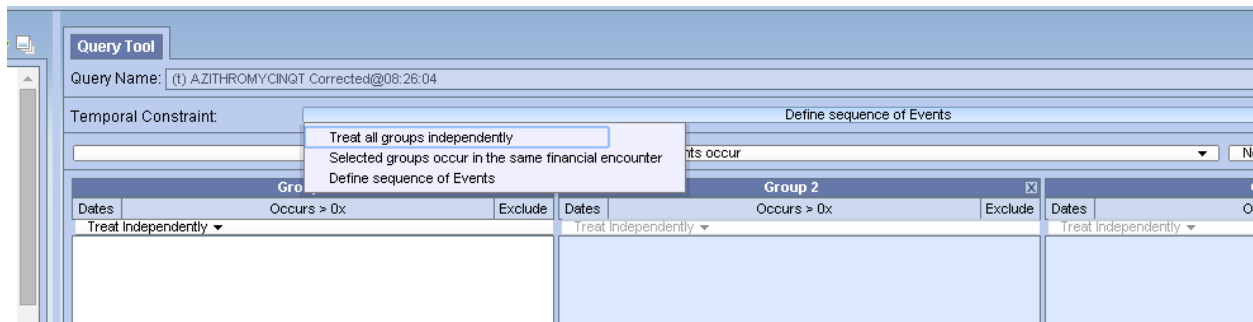
The screenshot shows the i2b2 Query & Analysis Tool interface. On the left, there is a 'Navigate Terms' panel with a tree view of medical terms. Below it is a 'Workplace' panel showing a list of queries. The main area is the 'Query Tool' where a query named 'QT Connected@08:22:14' is displayed. The query is defined by a temporal constraint: 'QT Connected >= 480 MS'. The results panel at the bottom shows 'Number of patients' as 77518 for the query 'QT Connected@08:22:14'.

Given that we now know our base-rate of patients (77518) we want to explore the number of patients who have ever been prescribed azithromycin. Here we pull over the second query in the demo, “Number of patients with QTC \geq 480 and on azithromycin ever”. Your screen should look as follows:

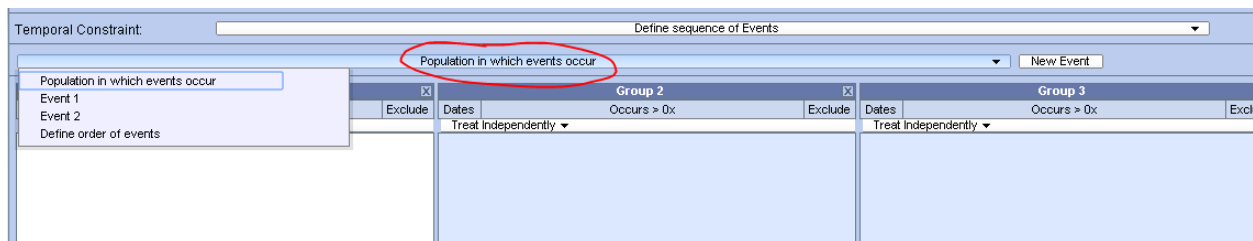
The screenshot shows the i2b2 Query & Analysis Tool interface with the second query. The 'Query Tool' panel now shows a query named 'QT Conn Azithr@08:23:59'. The query is defined by two temporal constraints: 'QT Connected >= 480 MS' and 'AZITHROMYCIN'. The results panel at the bottom shows 'Number of patients' as 4160 for the query 'QT Conn Azithr@08:23:59'.

However, what if we want to know who has had a long QT following the prescription to azithromycin? We use the *Temporal Constraint* function at the top center of the interface, circled in red.

Once you click on that field you should see the following then you'll select *Define sequence of Events*.



Next you'll change the query from a *population in which events occur* and start adding the events of interest.



You'll notice the option for *Define order of events* option. Once you have each event defined like you've done in the past with a simple query, you'll decide how you want to order events. I've skipped over the events space since it looks and behaves exactly like the environment you've seen in the past, you're just simply defining the criteria of each event.

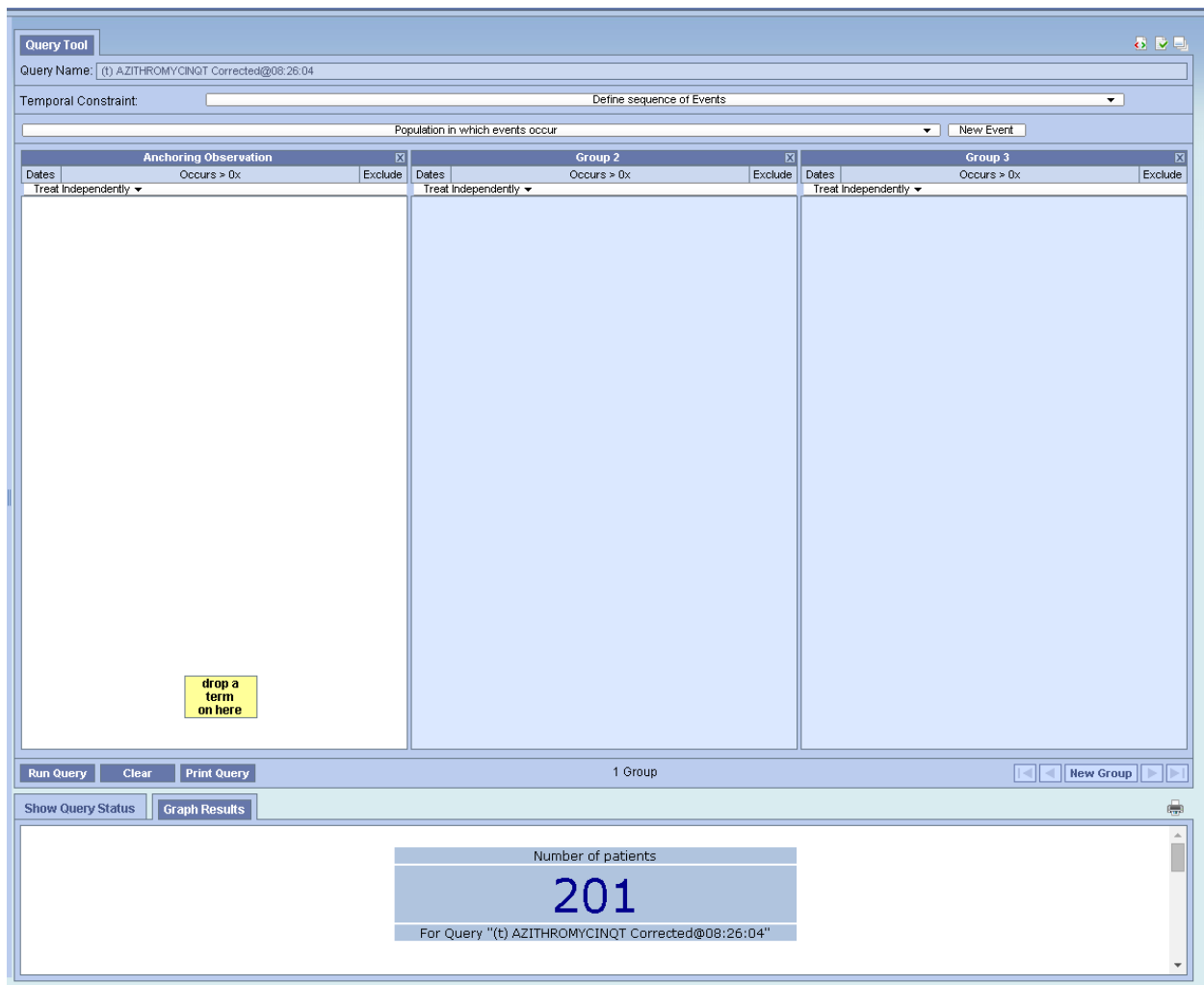
Now in the *Define order of events* interface you'll see some new options. Your screen should look as the one below.

The screenshot shows a software window titled "Query Tool". At the top, there is a "Query Name" field containing "(1) AZITHROMYCINQT_Corrected@08:26:04". Below this, there are two tabs: "Temporal Constraint" (which is active) and "Define sequence of Events". Under the "Temporal Constraint" tab, there is a "Define order of events" dropdown menu and a "New Event" button. The main area of the window contains two event definitions. The first event, "Event 1", is defined with "Start of" set to "the First Ever" and "Occurs Before" set to "Event 2". The second event, "Event 2", is defined with "Start of" set to "the First Ever". Below these, there are two checkboxes: "By" (checked) and "And" (unchecked). The "By" checkbox is followed by a dropdown menu set to "≤", a text input field containing "60", and a dropdown menu set to "day(s)". The "And" checkbox is followed by a dropdown menu set to "≤", a text input field containing "1", and a dropdown menu set to "day(s)".

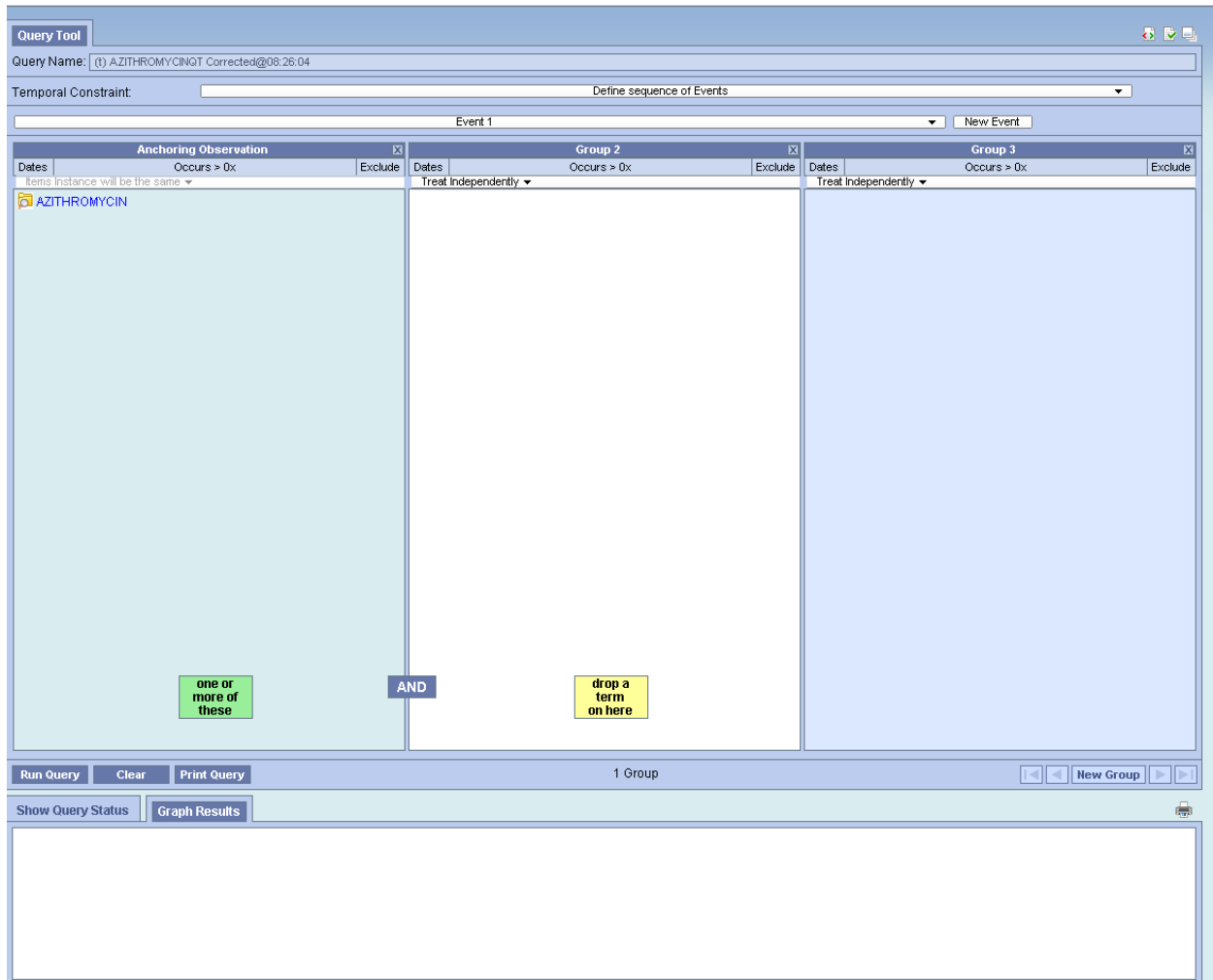
Here you'll see that we've said that persons falling into the first event (Event 1) will happen before (we defined this as patients with azithromycin) will be followed by the second event (Event 2) *BY* at least 60 days. You can choose to not include these kinds of restrictions, but for the sake of clinical interest we've included the restriction that patients should have a prescription of azithromycin at least 60 days prior to a QTC ≥ 480 .

Two important points about this query: we did not define a cut-off date, nor have we defined start-date. This means that we could find patients who have azithromycin prescribed any time before the long QTC and while it may seem natural to expand on your query at this point, it may push i2b2 beyond its intended purpose. Thus you may end up with very small results. To do these kinds of examinations, it would be best to contact the Regenstrief Data Core for these kinds of queries. Second, many investigators are interested in the outcomes of something due to an epoch effect—e.g., azithromycin was really popular in XXXX. This would also be a question best explored with the Regenstrief Data Core.

Now that we have explored the *Define sequence of events* function, let's review the output of this query. You'll notice that the query has already been generated as, "Number of patients with QTC ≥ 480 within 60 days after being on azithromycin". Once you pull this query over, your screen should look as the image below:



You'll notice that the pre-defined query here does not show you any concepts in any grouping. This is because you'll have to select which *Event* you're looking at. So for example, if we select Event 1 from the *Population in which occurs* bar at the top, you'll see the events. Select event 1, your screen should look like the image below:



Event two will look very similar, but will specify the conditions for Long QTC.

It is important to point out at this time, that time-series queries in i2b2 can be some of the most computationally intensive queries you can run therefore it is best to build on the previous methods if at all possible.

At this point let's explore a final IUH query, that is unexpected correlations.

Task 3: Hearing loss and Testicular Cancer

In demo section 3, you'll only see two queries: "Number of patients with testicular cancer", and "number of patients with testicular cancer and hearing loss". This is one of the principle use cases for i2b2 capability because it is an unexpected result to those not familiar with the phenomenon and it demonstrates that while some small-population research is not best conducted through i2b2. In the first query, we are looking for the testicular cancer icd-9's. As before, drag over the first concept labeled, "Number of patients with testicular cancer (ICD9)", your screen should look like the image below:

Query Tool

Query Name: ICD9:186 Malign@08:15:20

Temporal Constraint: Treat Independently

Group 1	Group 2	Group 3
<div>Dates</div> <div>Treat Independently</div> <div>Occurs > 0x</div> <div>Exclude</div> <div>ICD9:186 Malignant neoplasm of testis</div>	<div>Dates</div> <div>Treat Independently</div> <div>Occurs > 0x</div> <div>Exclude</div>	<div>Dates</div> <div>Treat Independently</div> <div>Occurs > 0x</div> <div>Exclude</div>

one or more of these

AND

drop a term on here

Run Query Clear Print Query 1 Group New Group

Show Query Status Graph Results

Number of patients

8956

For Query "ICD9:186 Malign@08:15:20"

However when we combine testicular cancer with the icd-9 for *hearing loss* (the second query), we find the image below:

Query Tool

Query Name: ICD9:18-ICD9:38@08:16:42

Temporal Constraint: Treat Independently

Group 1	Group 2	Group 3
Dates Treat Independently Occurs > 0x Exclude ICD9:186 Malignant neoplasm of testis	Dates Treat Independently Occurs > 0x Exclude ICD9:389 Hearing loss	Dates Treat Independently Occurs > 0x Exclude

one or more of these AND one or more of these AND drop a term on here

Run Query Clear Print Query 2 Groups New Group

Show Query Status Graph Results

Number of patients

153

For Query "ICD9:18-ICD9:38@08:16:42"

This suggests that IUH has seen at least 153 patients who have had testicular cancer AND have had a diagnosis of hearing loss—suggestive of hearing loss because of Cisplatin—however, this is an illustrative point where investigators should reach out to the Regenstrief Data Core. The Regenstrief Data Core can confirm feasibility of patient populations using the INPC (Indiana Network for Patient Care) and other tools which search text documents not stored otherwise.

Version Guide

Version Number	Author	Date
1	Nate Ring	9-30-2015