



Clinical & Translational Science Institute
of Southeast Wisconsin

USER GUIDE

Clinical Research Data Warehouse

Research Data Tools

i2b2 & TriNetX query tools help research teams discover and validate patient cohorts using existing well-aggregated, de-identified clinical data from various sources.
Honest Broker is an IRB-approved research data extraction tool.

CTSI Biomedical Informatics Team
CRDW@mcw.edu

Table of Contents

Clinical Research Data Warehouse (CRDW)	3
CRDW Tools	3
Data Sources (start dates)	3
Data Integrity Partners	4
Data Access Agreements/Requirements.....	4
Data Team/Support.....	4
CRDW Content for Cohort Query and Data Extraction.....	5
Data Querying	6
i2b2 – Informatics for Integrating Biology and the Bedside.....	6
<i>Logging into i2b2</i>	6
<i>Introduction</i>	7
<i>Building a Query</i>	10
<i>Running a Query</i>	12
<i>Using Advanced Concept Search Techniques</i>	16
<i>Practice: Basic Querying with i2b2</i>	20
<i>Redefining a Previous Query</i>	24
<i>Adding More Query Groups</i>	24
<i>Using Temporal Constraints</i>	26
<i>Analysis Tools</i>	29
TriNetX	33
Data Extraction	34
Honest Broker	34
<i>Downloading De-Identified Data</i>	35
<i>Requesting Identified Data</i>	39
EMERSE Text Search Tool	40
FAQs	41
<i>What should I do if the i2b2 screen is blank when I log in?</i>	41
<i>Can I store de-identified data extracts from Honest Broker on an encrypted flash or thumb drive?</i>	41

Clinical Research Data Warehouse (CRDW)

The CRDW is an aggregation of clinical and demographic data from various sources.

- **De-identified** patient data can be accessed through query tools and extracted using the Honest Broker tools
- **Identified** patient data can be extracted by researchers with an approved IRB protocol using the Honest Broker tools and process

CRDW Tools

All data warehouse tools are available at <https://ctsi.mcw.edu/ctri/resources/bmi-links/>

Data Warehouses

- Froedtert & The Medical College of Wisconsin
- Children's Hospital of Wisconsin (Children's Specialty Group data only)

Query Tools

- i2b2 – Informatics for Integrating Biology and the Bedside
- TriNetX (<https://www.trinetx.com/trinetx-live/>)

Data Extraction Tools

- Honest Broker (HIPAA-compliant, IRB-approved)
- Jupyter Hub (double encrypted big-data environment)

Data Sources (start dates)

- **Epic EHR (electronic health record) Systems**
 - Froedtert & Medical College of Wisconsin (October 2004)
 - Froedtert Menomonee Falls Hospital (May 2012)
 - Community Physicians Clinics (July 2013)
 - Froedtert West Bend Hospital (September 2013)
 - Children's Wisconsin (November 2013)
 - Children's Specialty Group only
 - Children's Medical Group data not available
- **OnCore**
 - Biospecimen data from MCW tissue bank (2012)
 - Clinical Trials enrollment (yes/no)
- **NAACCR Tumor Registry (SJH – 2001, CMH - 1989, FMLH – 1989)**
- **GE/IDX Physician Billing System (~1999)**
 - Group 3 Medical College Physicians
 - Group 4 Children's Specialty Group
 - Group 7 FMCCP (Community Physicians)
- **Froedtert Hospital legacy systems retired in 2012**
 - Affinity (clinical ordering and billing system)

Data Integrity Partners

- MCW Institutional Review Board (IRB)
- Froedtert Office of Clinical Research and Innovative Care Compliance (OCRICC)
- F&MCW Epic Core Team, Clinical Informatics & Pharmacy Informatics

Data Access Agreements/Requirements

Our access form is available at <https://ctsi.mcw.edu/ctri/cda/crdw/>

- Non-MCW faculty must have an MCW faculty sponsor to apply for access
- We request one access agreement per project when new team members need access
- If you are in the early stages of project development, you may be general in your description of desired data and intended use of our tools.
- Form pages include:
 - Legal Agreement (faculty PI/sponsor info must be entered on this page)
 - Research Project
 - Project Team
 - Cost Structure Notice Disclaimers

CRDW users must have current CITI training in Human Subjects Research

- The Biomedical Learner track is preferred

CRDW users must have an MCW network account

- Faculty and staff from CTSI member organizations may access CRDW
 - Children's Wisconsin
 - Versiti
 - Milwaukee VA
 - Marquette University
 - MSOE University
 - University of Wisconsin – Milwaukee
 - Froedtert Health
- Biomedical Informatics will request MCW network accounts for non-MCW users

Data Team/Support

Technical Issues

- i2b2 or Honest Broker, contact CRDW@mcw.edu
- TriNetX, contact support@trinetx.com

User Access & Training Contacts

- Chelsea Spangenberg, CTSI Program Manager
 - cspangenberg@mcw.edu
- Kris Osinski, CTSI Biomedical Informatics Business Analyst
 - kosinski@mcw.edu

CRDW Content for Cohort Query and Data Extraction

Data Domain	Source(s)	i2b2	TriNetX	Honest Broker
Allergy History	Epic	Yes	No	Yes
Biospecimens in MCW Tissue Bank	OnCore	Yes	No	Yes
Clinical Trial (enrolled Y/N)	OnCore CTMS	Yes	No	No
Contact Information (identified only)	Epic	No	No	Yes
Demographics	Epic, SSDMF	Yes	Yes	Yes
Diagnoses (ICD-9/ICD-10)	Epic, GEIDX billing	Yes	Yes	Yes
Diagnostics Results <ul style="list-style-type: none"> ▪ Clinical Labs ▪ Pulmonary Function Testing ▪ Cardiovascular Testing ▪ Radiation Oncology Treatment 	Epic	Yes	Yes	Yes
Encounters/Visit Details	Epic, GEIDX billing	Yes	Yes	Yes
Encounters – Future (identified only)	Epic	No	No	Yes
Genomics (Ambry, Foundation, Invitae, Tempus + Summary Table)	Various	Yes	Yes	Yes
Imaging Orders	Epic	No	No	Yes
Immunization	Epic	Yes	Some	Yes
Medications Administered	Epic	Yes	No	Yes
Medications Ordered	Epic	Yes	Yes	Yes
NAACCR (FMLH, FHMF, and FHWB)	CNExT Tumor Registry	Yes	Some	Some
OB/GYN Mother & Baby	Epic	No	No	Yes
Problem List	Epic	Yes	No	Yes
Procedures <ul style="list-style-type: none"> • Epic Orders • CPT/HCPCS • ICD (Inpatient hospital billing) 	Epic, GEIDX billing	Yes	Yes	Yes
Providers	Epic	Yes	No	Some
Social History/Determinants	Epic	Yes	Some	Yes
Surgical Case	Epic	No	No	Yes
Vitals (outpatient encounters only)	Epic	Yes	Yes	Yes

Domains not on this list may be available by custom extract from the CTSI Biomedical Informatics core on a fee-for-service basis.

Data Querying

i2b2 – Informatics for Integrating Biology and the Bedside

i2b2 is an informatics framework designed to make de-identified clinical data accessible for project feasibility and cohort discovery. The **i2b2** tool offers “drag-and-drop” functionality for selecting query criteria and generating cohort counts and patient sets.

Logging into i2b2

i2b2 is compatible with many web browsers include **Firefox, Chrome & Safari**

Both **i2b2** tools are accessible at <https://ctsi.mcw.edu/ctri/resources/bmi-links/>

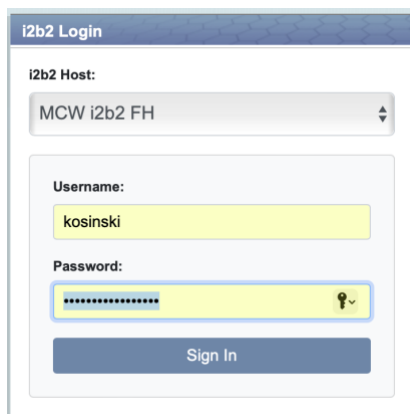
Froedtert Health System i2b2

- Includes all Froedtert & Medical College of Wisconsin hospitals and clinics
- Current through 2/17/2022

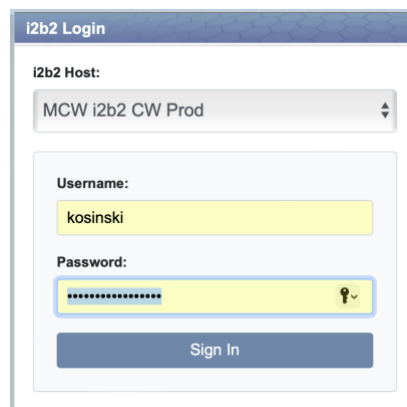
Children’s Wisconsin i2b2

- Includes data for Children’s Specialty Group
- Current through 1/31/2022
- *NOTE: The Children’s data warehouse DOES NOT include data from Children’s Medical Group*

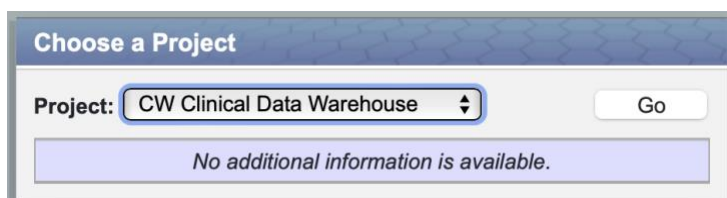
Log in with your **MCW network** username and password, click **Sign In**, then click **Go**



The screenshot shows the 'i2b2 Login' form for 'MCW i2b2 FH'. It features a dropdown menu for 'i2b2 Host' with 'MCW i2b2 FH' selected. Below are input fields for 'Username' (containing 'kosinski') and 'Password' (masked with dots). A 'Sign In' button is at the bottom.



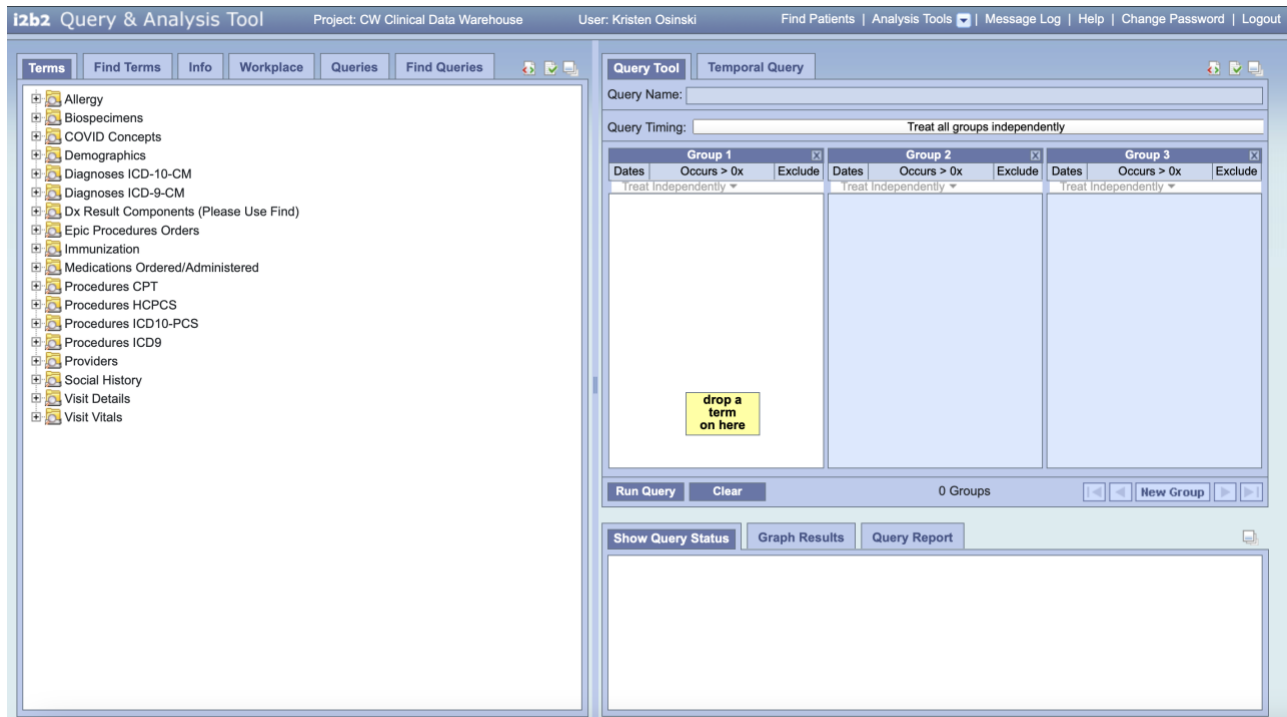
The screenshot shows the 'i2b2 Login' form for 'MCW i2b2 CW Prod'. It features a dropdown menu for 'i2b2 Host' with 'MCW i2b2 CW Prod' selected. Below are input fields for 'Username' (containing 'kosinski') and 'Password' (masked with dots). A 'Sign In' button is at the bottom.



The screenshot shows the 'Choose a Project' form. It has a dropdown menu for 'Project' with 'CW Clinical Data Warehouse' selected. A 'Go' button is to the right. Below the form is a message box that says 'No additional information is available.'

Introduction

The **i2b2 Query & Analysis Tool** screen should appear upon logging in



Tabbed Panes

The **i2b2 Query & Analysis Tool** consists of many panes and tabs for searching, building, running and storing queries. The tabbed panes include:

1. **Terms** – a list of concept groups in a hierarchical folder structure referred to as the Terms Tree or Concept Tree
2. **Find Terms** - contains useful search functions that help locate query concepts within the tree utilizing words and phrases or specific codes to describe the query concept
3. **Info** – provides details information for a term highlighted in Terms or Find Terms
4. **Workplace** – a shared directory area used to store and share queries for later review. Each user has a personal query folder named by their login ID.
5. **Queries** – a personal library of previous queries and results for each user
6. **Find Queries** - a search functions to locate previously saved queries.
7. **Query Tool** – the work area used to collect query terms/concepts, define constraints and build the logical relationships between query concepts/terms
8. **Temporal Query** – the work area used to build temporal relationships between groups of terms/concepts to help refine a cohort definition

- 9. **Show Query Status** – the area used to display the progress of a running query. When finished, query results will display in the **Graph Results** and **Query Report** tabs.

Tabbed Pane Features

Several of the **Tabbed Panes** provide additional features that can help control and view concepts or results within that pane.



The **Show Options** icon opens a pane-specific selection box which enables a user to control how content may be displayed or used within the selected pane.



The **Resize Workspace** icon is available in all i2b2 panes. This feature enables toggling between the condensed pane size and a vertically expanded view of the pane which helps reduce the need for vertical scrolling.

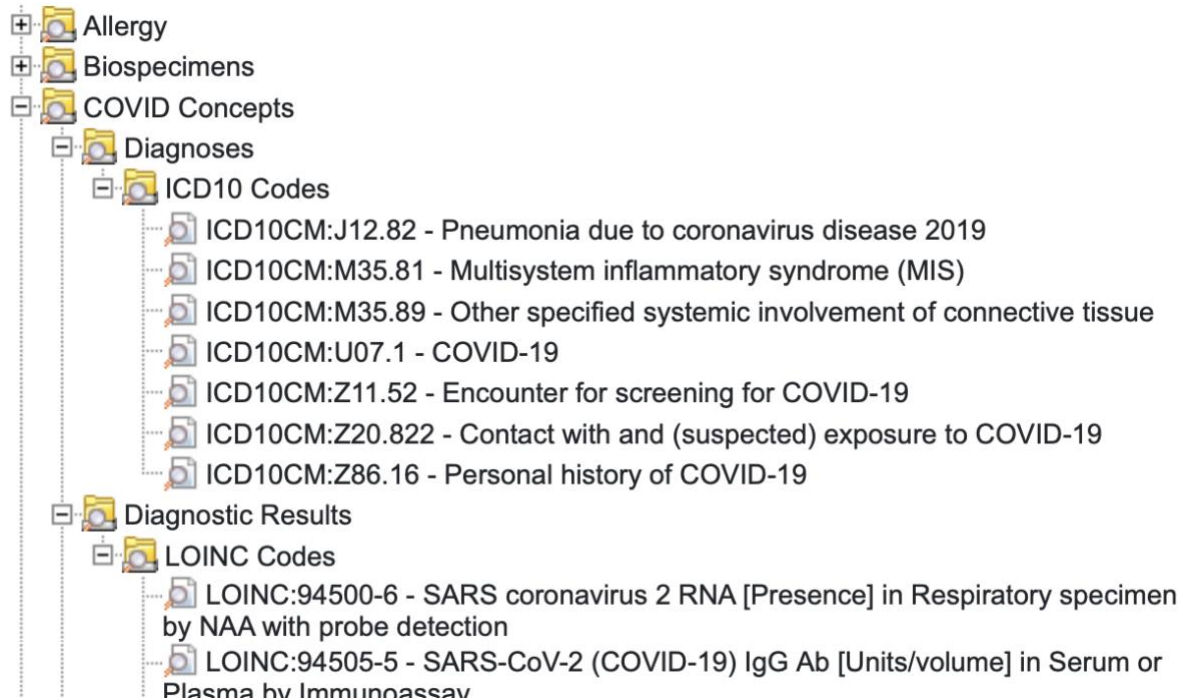


The **Refresh** icon is available in the panes which display running and finished queries.

Concept Tree

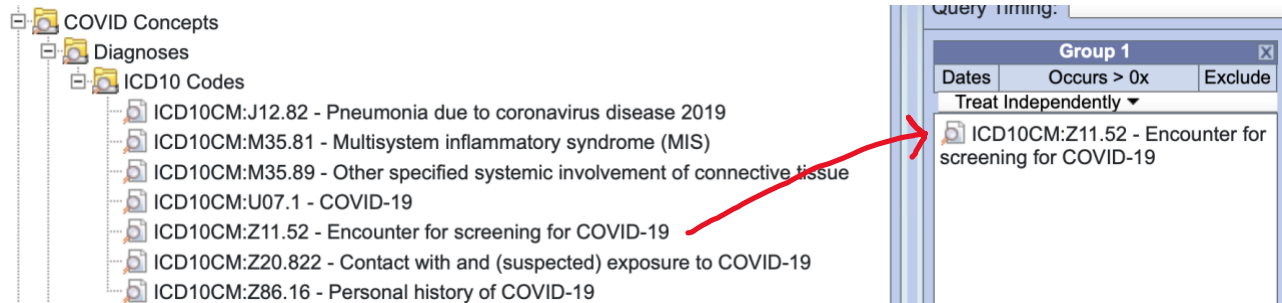
The **Terms** pane provides a hierarchical tree of concepts that can be used as query criteria.

Click the [+] and [-] icons to expand and contract sections within the Concept Tree.



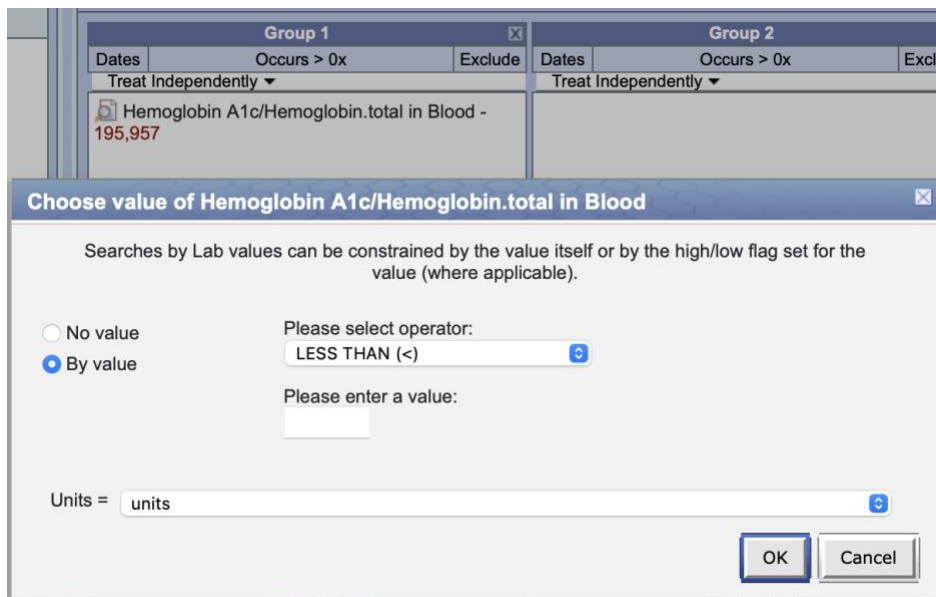
Using Drag & Drop

When a desired query concept is found... click, hold and **Drag** the concept from the **Navigate Terms** pane and then **Drop** it into the next available **Query Tool - Group** panel. Drag & Drop automatically adds the concept to the query.



Some concepts are enabled with “modifier” options which require users to choose more specific parameters for the criterion. These options will be presented automatically when trying to drop a concept into a **Group** panel.

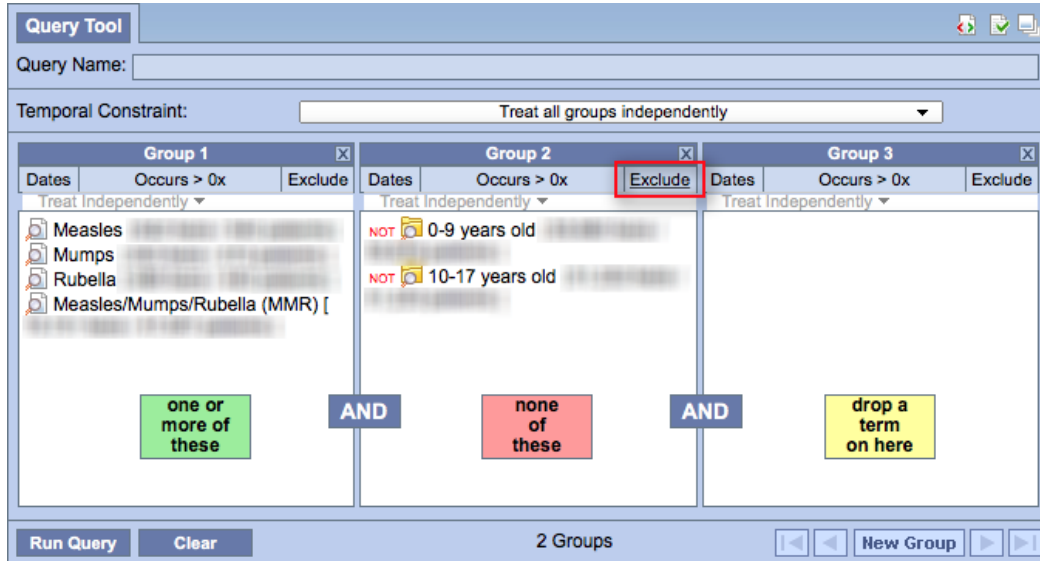
For example, when you drag & drop a Diagnostic Result (Labs) concept into a **Group**, the following pop-up appears:



- To find all patients with resulted tests regardless of the result values, click **No Value**
- To set a specific value range, select **By Value**, select an operator, and then enter the appropriate value range for the concept
- Click **OK** to set the criteria and complete the drop

Building a Query

Multiple concepts can be used to create complex query criteria and refine a potential patient cohort using the Drag & Drop tools of i2b2.

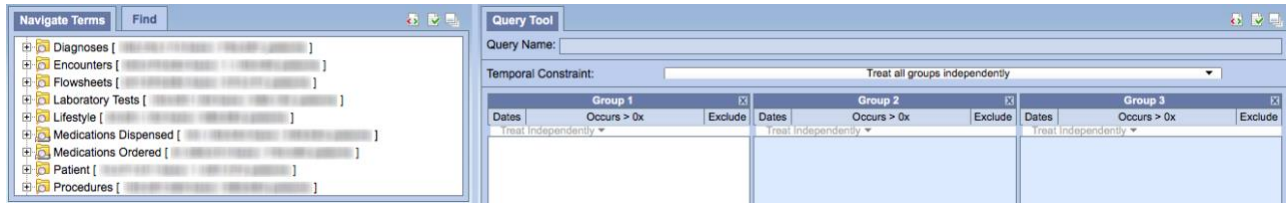


- When a concept is dragged into a **Group** box, a green message box (*one or more of these*) is displayed to indicate that all concepts in that Group will be considered within the query.
- Concepts within a single panel will be connected by 'OR' logic, meaning the population is broadened to include all patients who fit any of those criteria.
- To exclude concept(s) in a **Group** from the query, click the **Exclude** button in the upper right corner of the Group. A red message box (*none of these*) is displayed to indicate that none of the concepts in that Group will be considered within the query.
- Concepts between panels will be connected by 'AND' logic, meaning the population is narrowed to include only patients who fulfill all requirements.

Rule of thumb: put the most specific criteria set in **Group 1** to narrow the cohort and then broaden the specificity as criteria groups are added. This method decreases the amount of indexing needed behind the scenes and reduces query processing time.

For example: we want to search for **adult women** diagnosed with **schizophrenia**.

To create this query, we will start in the **Terms** tab and select concepts from the **Patient** and **Diagnoses** groups. Click the [+] icon next to a concept group to open it.



We will Drag & Drop the concepts into the **Query Tool** and then run the query.

The concept “schizophrenia” is most specific and can be found within the **Diagnoses** concept group under the path *Mental Disorders/Psychosis/Schizophrenic Disorders*. Because we want to include all possible schizophrenic disorders, we drag the entire folder to **Group 1**. Before it will drop in, though, we must indicate whether to include all diagnosis sources in our query or just certain ones. For this example, we choose **No Value** to include all diagnosis data sources.

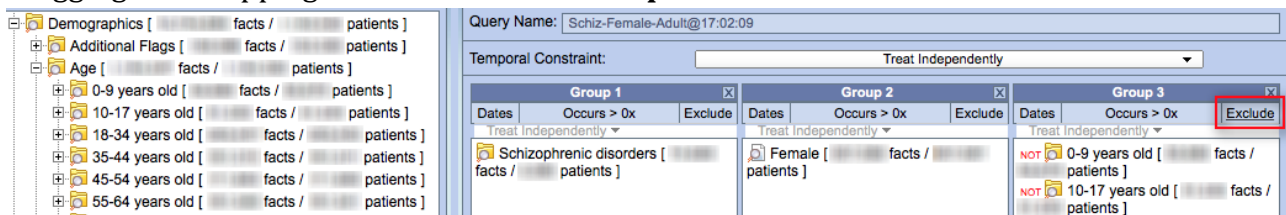
The concept “women” is found within the **Patient** concept group under the hierarchy path *Demographics/Gender*. Select *Female*, and then drag and drop the concept into **Group 2**.

The concept “adult” is the least specific of the three criteria and is also found within the **Patient** concept group under *Demographics/Age*.

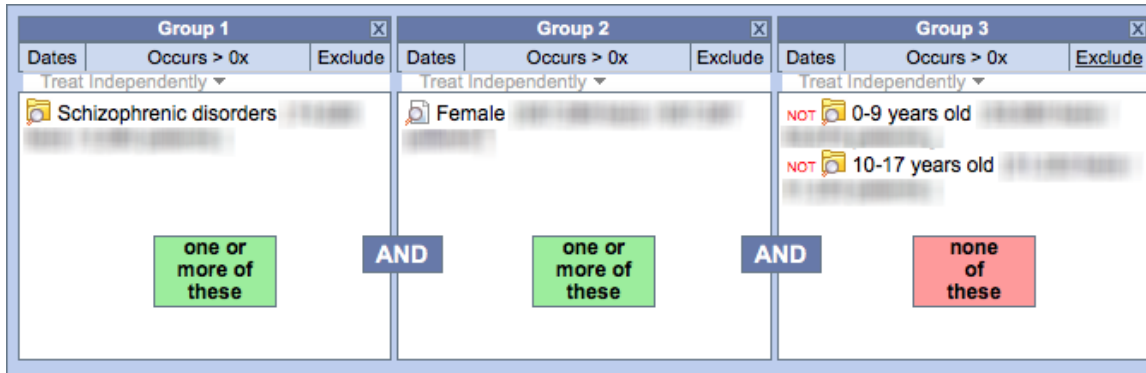
There are two possible approaches to defining this criterion:

- Include all age subgroups containing age 18 and greater, or
- Exclude all age subgroups under age 18.
-

The more efficient approach is to exclude the two “less than 18” age subgroups vs. separately dragging and dropping all the others into **Group 3**.



One by one, select the concept folders for **0-9 years old** and **10-17 years old** and drag each one into **Group 3**. Then click the **Exclude** button in **Group 3** to remove patients in these two age groups from the query. The red message box (*none of these*) appears in **Group 3**.



Logically, our query for **adult women** with **schizophrenia** is organized as follows: Patients who have *Schizophrenic disorders* AND are *Female* AND Age NOT (*0-9 years old* OR *10-17 years old*)

Running a Query

Once the query criteria are confirmed, click the **Run Query** button to name the query, assign the query result type(s), and run the query. Each unique query should have a distinct name. The default name is generally an abbreviation of the query criteria groups followed by the current time, but the query name can and should be customized for good query management.

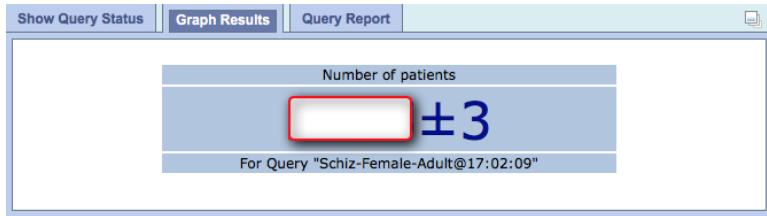
By default, the query result type is **Number of patients** which provides a distinct patient count. *Patient set* and *Timeline* options are discussed in the [Analysis Tools](#) section of this document. Do not use the DataBuilder query option.

The 'Run Query' dialog box contains the following elements:

- A text input field for the query name, containing 'Test Query@12:15:22'.
- A section titled 'Please check the query result type(s):' with four radio button options:
 - Patient set
 - Number of patients
 - Timeline
 - DataBuilder query
- 'OK' and 'Cancel' buttons at the bottom.

Click **OK** to begin running the query.

The **Graph Results** tab displays the query's progress as it processes as well as a summary of the results when it is complete.



Note: Query results are limited to aggregate patient counts which are further obfuscated (obscured) by randomly returning record counts with a deviation of ± 3 records. Results of less than 11 are never displayed.

The **Show Query Status** tab also displays the aggregate patient count as well as the number of seconds it took to run the query.



The **Query Report** tab provides a printable report which includes details about the query process, all criteria included in the query definition, and a summary of the query results.

Show Query Status | Graph Results | Query Report


i2b2 Query Report Print Report

The query entitled "Schiz-Female-Adult@17:02:09" submitted on 2016-05-09 17:02:21, was successfully completed on 2016-05-09 17:02:21. This query was performed by "mcwcorp\ [redacted]". The search was completed in 0 seconds.

Query Definition

Temporal Constraint: Treat All Groups Independently

All Groups

 **Schizophrenic disorders** [10,000 facts / 10,000 patients]


Diagnoses \ Mental Disorders \ Psychosis \ Schizophrenic disorders

Independent of Visit

From earliest date available to latest date available

of times an item is recorded is > 0

AND

 **Female** [100,000 facts / 100,000 patients]


Demographic \ Gender \ Female

Independent of Visit

From earliest date available to latest date available

of times an item is recorded is > 0

AND NOT

 **0-9 years old** [10,000 facts / 10,000 patients]


Demographic \ Age \ 0-9 years old

Independent of Visit

From earliest date available to latest date available

of times an item is recorded is > 0

OR

 **10-17 years old** [10,000 facts / 10,000 patients]

Demographic \ Age \ 10-17 years old

Independent of Visit

From earliest date available to latest date available

of times an item is recorded is > 0

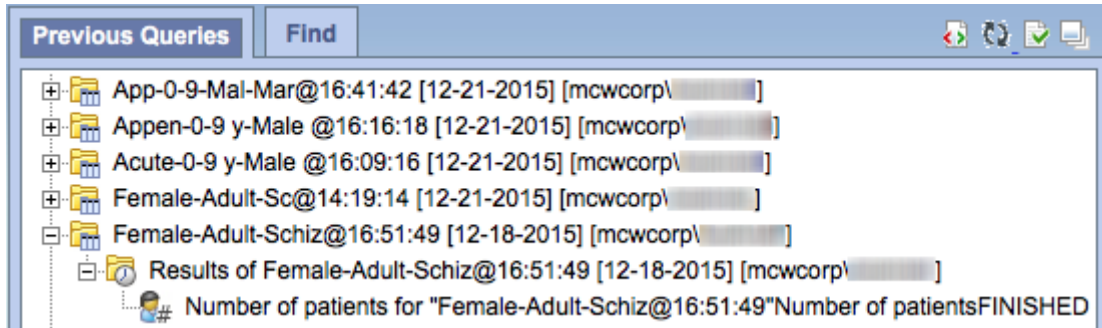
Query Results

Total Number of Cases

Total Patients Matching Query
[redacted] ±3

Print Report

Completed queries appear in the **Previous Queries** pane using a folder hierarchy display.



Note: If the query takes longer than 3 minutes, it will run in the background and will be noted as **PROCESSING** in the **Previous Queries** section until the query status is **FINISHED**.

The **Find** tab allows users to search within their own query list utilizing one or all of the following categories:

- Previous Query Name
- Previous Result Type
- Patient Number

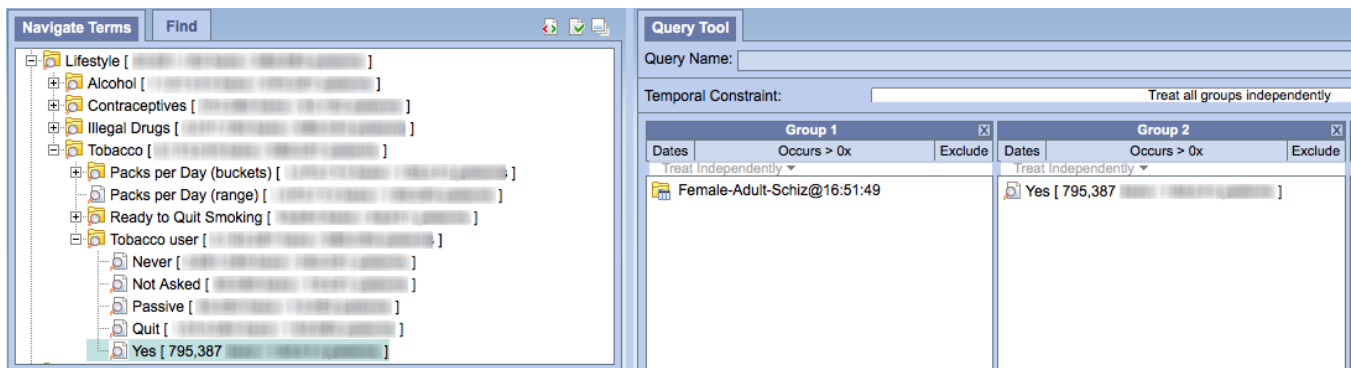
The default search parameter is *Containing*, but *Exact*, *Starting with* and *Ending with* options are also available.

A **Previous Query** can be used as criteria within a new query.

For example, our cohort of adult females with schizophrenic disorders can be dragged back into the Query Tool to serve as aggregated criteria for new potential cohorts such as:

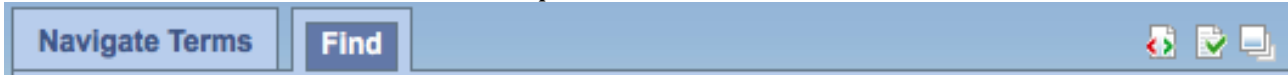
- Adult Female Schizophrenics who use tobacco (shown below)
- Adult Female Schizophrenics who have been diagnosed with diabetes
- Adult Female Schizophrenics who use tobacco and have been diagnosed with diabetes

Note: Tobacco use criteria is located in the **Lifestyle** concept group



Using Advanced Concept Search Techniques

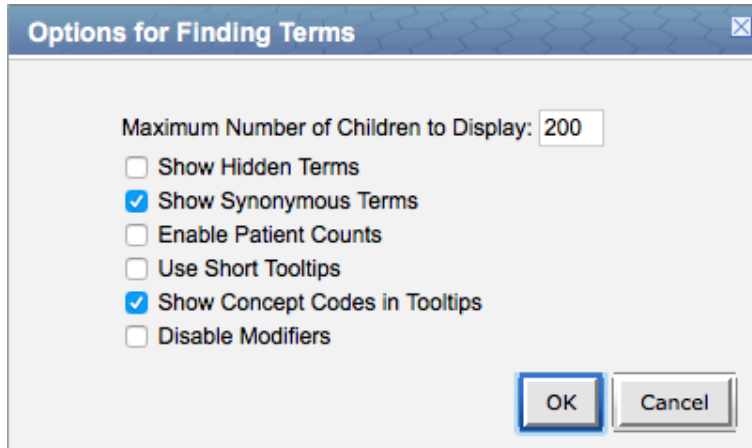
The **Find** tab offers two useful search options to find cohort criteria.



Find Tab Settings

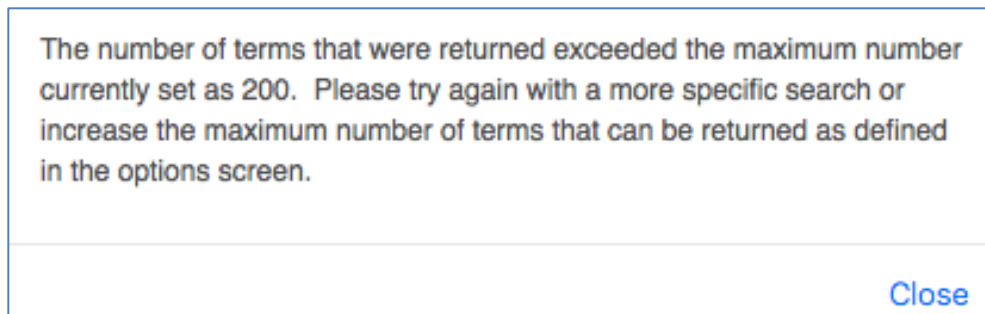


Click the **Show Options** icon to display the additional search functions box.



Maximum Number of Children to Display

If a search returns more than the maximum value set in this box, i2b2 will generate the pop-up message below. The default setting is 200, but you may increase this number as needed.



Show Hidden or Synonymous Terms

- Do not check the Show Hidden Terms box as we do not utilize hidden terms.
- **We recommend that users uncheck the Show Synonymous Terms option.** This option allows users to find concepts with alternate terms, but the search may return redundant results which could create redundant query criteria which, in turn, may inadvertently increase query processing time. Example: A search for “asthma” in Diagnoses returns 63 results with synonyms on vs. 22 with synonyms off.

Enable Patient Counts

Leave this option unchecked as we have already provided counts in concept display names.

Use Short Tooltips

Leave this option unchecked as the short name is no different than the concept display name.

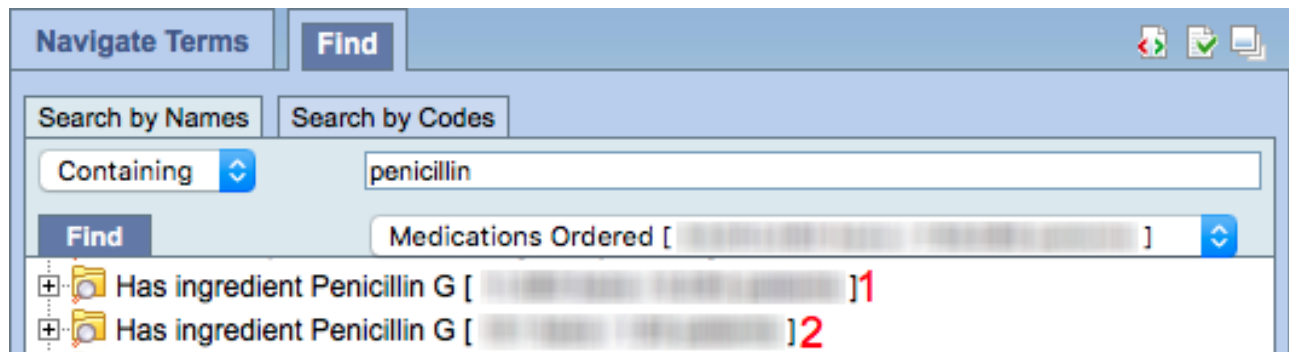
Show Concept Codes in Tooltips

Always keep this defaulted option checked. Tooltips are very useful in i2b2 because they show users the full path to each search result within the Navigate Terms concept tree.

Hover the cursor over each result to display its full path within the Concept Tree.

Example: A search for the term “penicillin” in Medications Ordered returns what appear to be redundant results, but their tooltips reveal two unique paths. Medication ingredients will frequently return multiple search results because a single ingredient may be contained within multiple pharmaceutical classes or subclasses.

Depending on the query requirements, it may be appropriate to include all, some or just one of the search results as query criteria.



Tooltip 1: `\Meds Ordered \ Penicillins \ Natural Penicillins \ Has ingredient Penicillin G - RXCUI:7980`

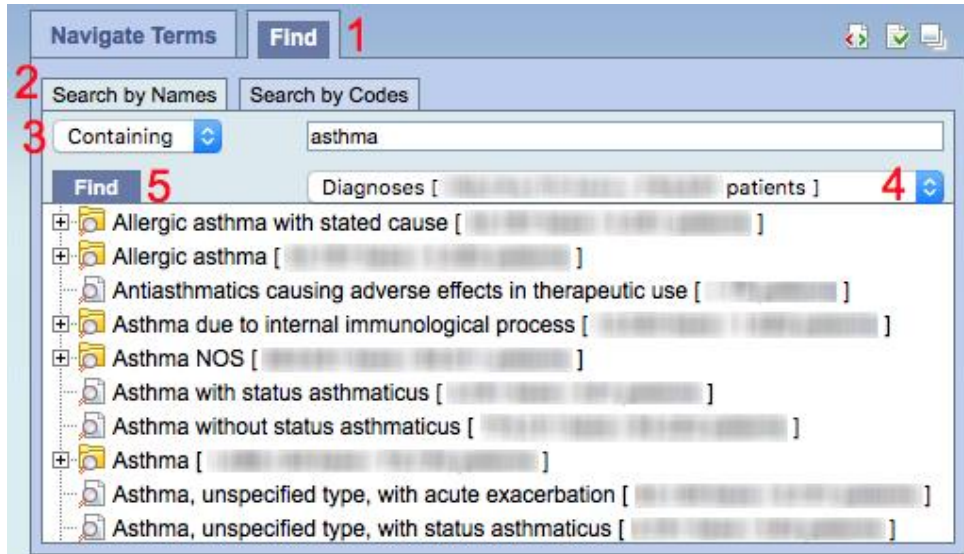
Tooltip 2: `\Meds Ordered \ Penicillins \ Penicillin Combinations \ Has ingredient Penicillin G - RXCUI:7980`

Disable Modifiers

Never check the disable modifiers option. Modifiers are critical to specifying query criteria.

Search by Names

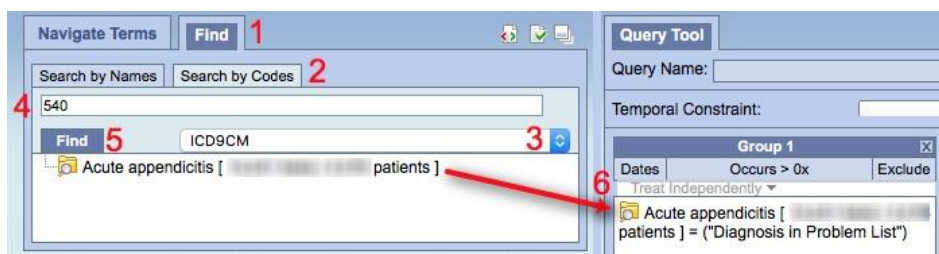
Search by Names offers significant flexibility in locating query criteria. Full words, partial words, or phrases can be used to locate pertinent query concepts.



1. Click the **Find** tab.
2. Click the **Search by Names** tab.
3. Click the **Containing** box to choose between *Containing*, *Exact*, *Starting with* or *Ending with* term search parameters, and then type the full or partial term in the text box to the right.
4. Click the **Any Category** box to view the available concept group filtering options. Select a specific concept group for a targeted search or select *Any Category* to search for a term within all concept groups.
5. Click the **Find** button. All terms matching the search criteria will be displayed in a list. Click the [+] icon next to a yellow concept folder to open it and search deeper for more specific concepts.

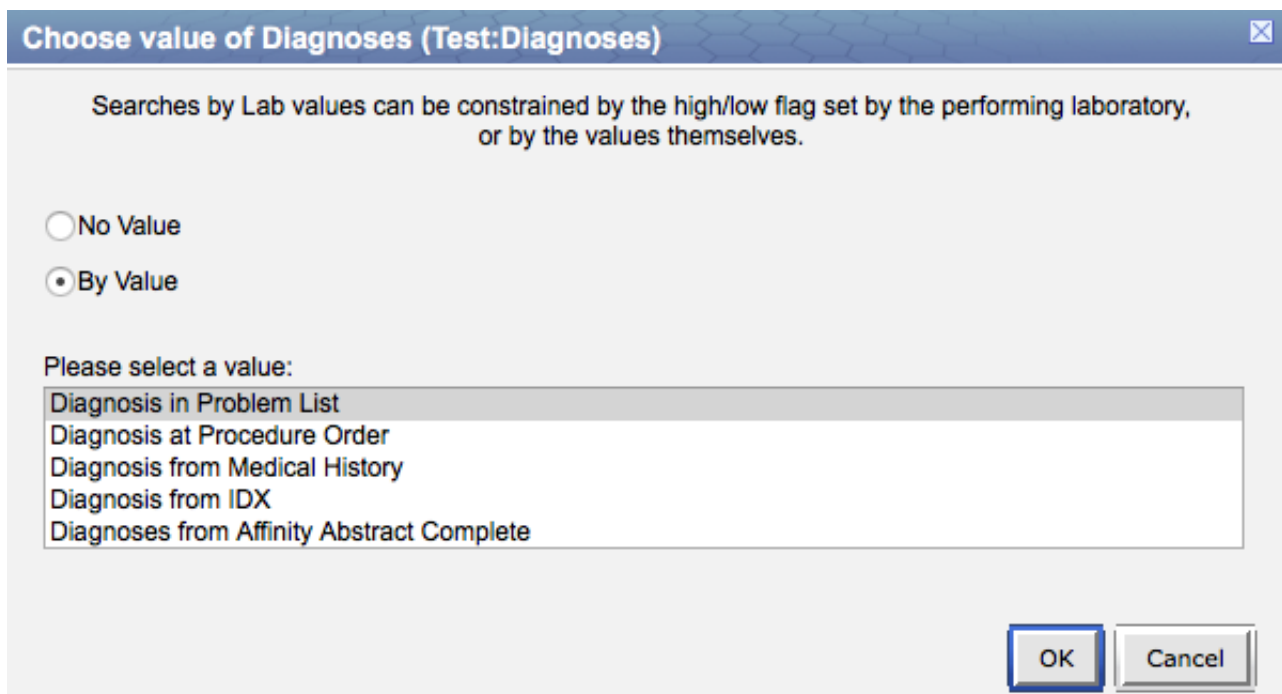
Search by Codes

In addition to searching by term names, **Find** also offers the ability to search within established clinical code sets such as *CPT*, *ICD-9* and *LOINC*. When users already know the codes they want to include in a query, this option allows for efficient, targeted searching.

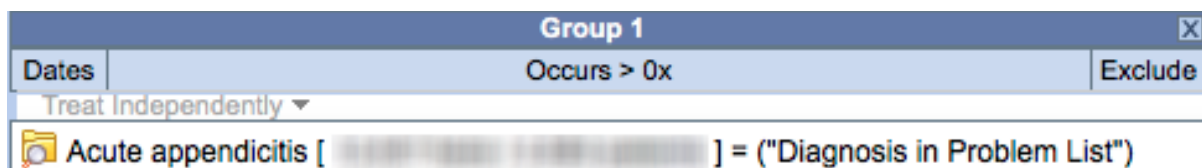


1. Click the **Find** tab.
2. Click the **Search by Codes** tab.
3. From the **Select a Coding System** drop down list, click on a code set (e.g., ICD9).
4. Type in a specific code (e.g., 540).
5. Click the **Find** button.
6. Drag the specific query concept into a Group within the **Query Tool** pane.

In this example, we elected to limit the query criteria to patients who have diagnosis code 540 in their Problem List by selecting 'Diagnosis in Problem List' within the By Value selection box.



When this type of constraint is placed on query criteria, it is literally displayed in quotes within the Group box after the concept display name with an '=' sign.



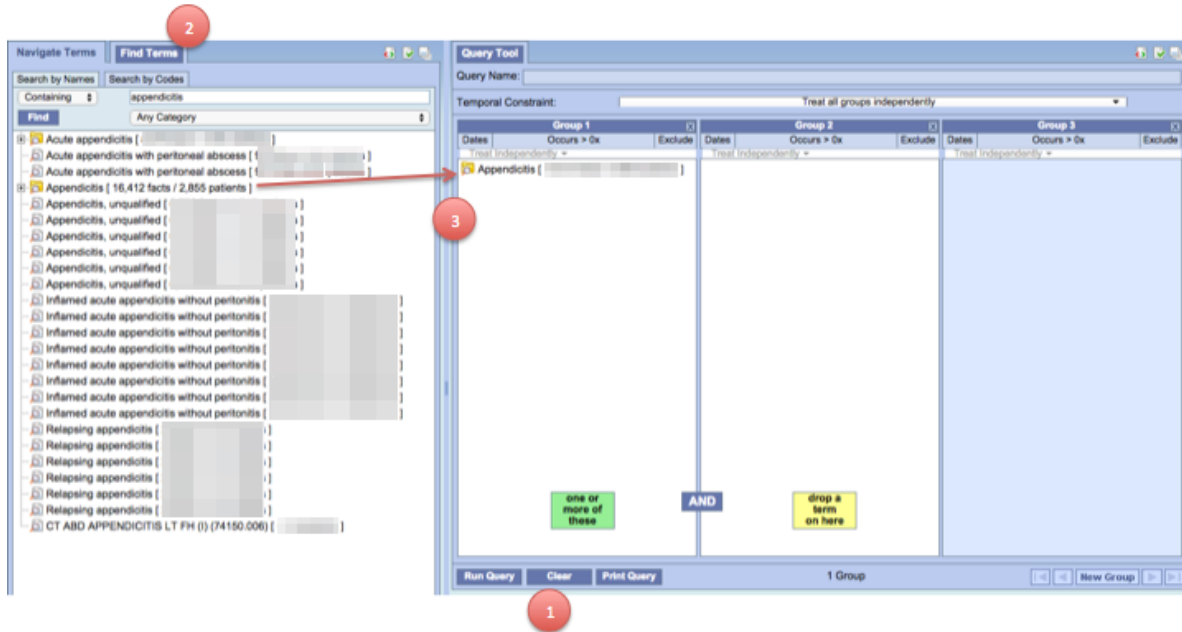
Practice: Basic Querying with i2b2

Identify a Patient Cohort

To illustrate how **i2b2** can be used for cohort discovery, let's try a sample scenario. We will build a query to identify:

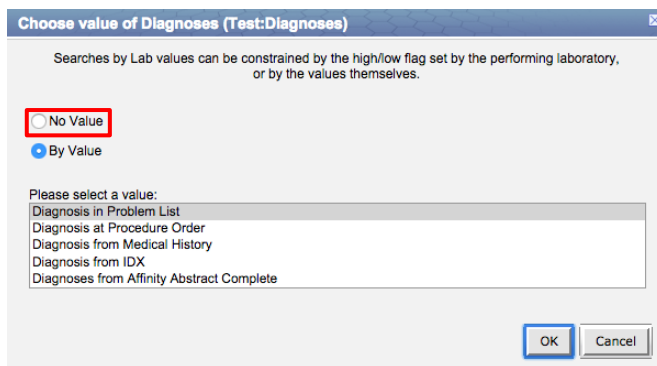
- Adult Male patients (age 18 and older)
- Diagnosed with Appendicitis during the period 1/1/2005 to the present.

Define the Diagnosis Criteria



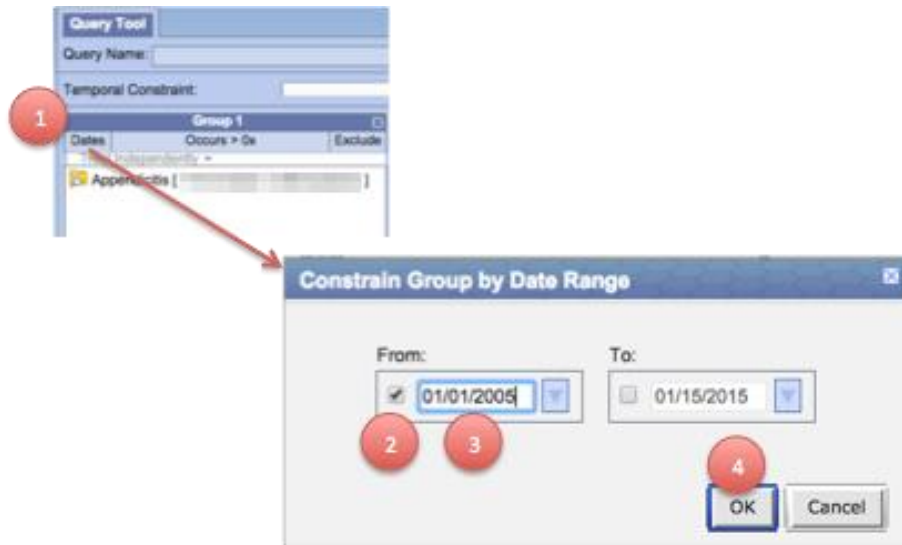
1. If existing query criteria are present in the **Query Tool**, click the **Clear** button to clear the Group box(es) and start a new query.
2. Click on the **Find** tab, and use **Search by Names** to find Appendicitis.
3. Click and hold on the Appendicitis folder to drag and drop it into **Group 1**.

Select “No Value” to choose diagnoses from any source.



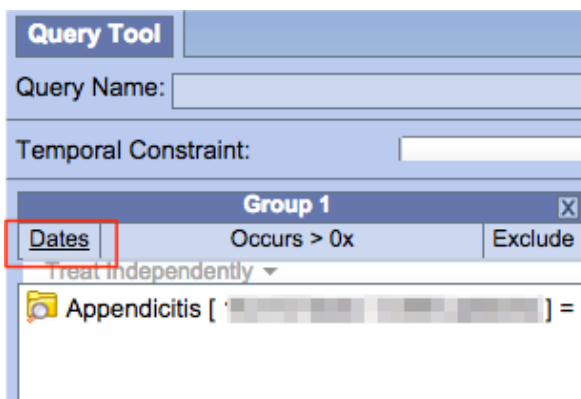
Define the Diagnosis Date Range Criteria

Use the **Dates** button inside the **Group 1** box to limit our Appendicitis diagnosis criteria to the specific date range defined for this cohort: 1/1/2005 to present.



1. Click on the **Dates** button.
2. Click the checkbox in the **From:** box.
3. Change the **From:** date to 1/1/2005. If the checkbox in the **To:** box is not selected, it will default to today's date. Click the checkbox in the **To:** box to customize the endpoint of the date range.
4. Click the **OK** button to set the criteria.

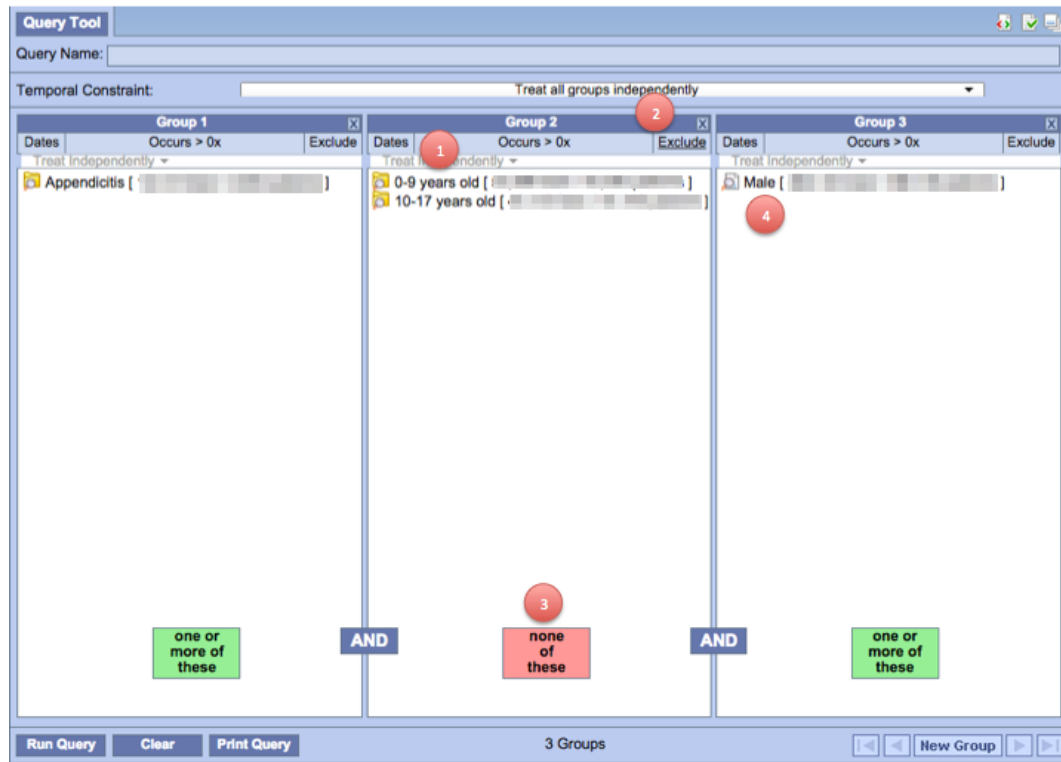
Note: Once a date range has been set within a Group, the **Dates** button will display as underlined (i.e. Dates).



Define the Demographic Criteria

Our patient cohort is limited to Adult Males.

Age and gender concepts are defined in **i2b2** as Patient Demographics.



1. First, go to **Navigate Terms** to open the Patient -> Demographics -> Age concept and then drag & drop both the *0-9 years old* and the *10-17 years old* folders into **Group 2** of the Query Tool.
2. Next, click the **Exclude** button inside the Group 2 box.
3. Notice that the red message box (*none of these*) appears within Group 2 to confirm that patients under the age of 18 will be excluded from the query.
4. Finally, go back to **Navigate Terms** to open the Patient -> Demographics -> Gender concept and then drag & drop the *Male* concept into **Group 3** of the Query Tool.

Before running the query, confirm that the entire query statement makes sense.

Our Adult Male Appendicitis query includes patients with:

Group 1: Diagnoses of *Appendicitis* diagnosed from *1/1/2005 to present*

AND

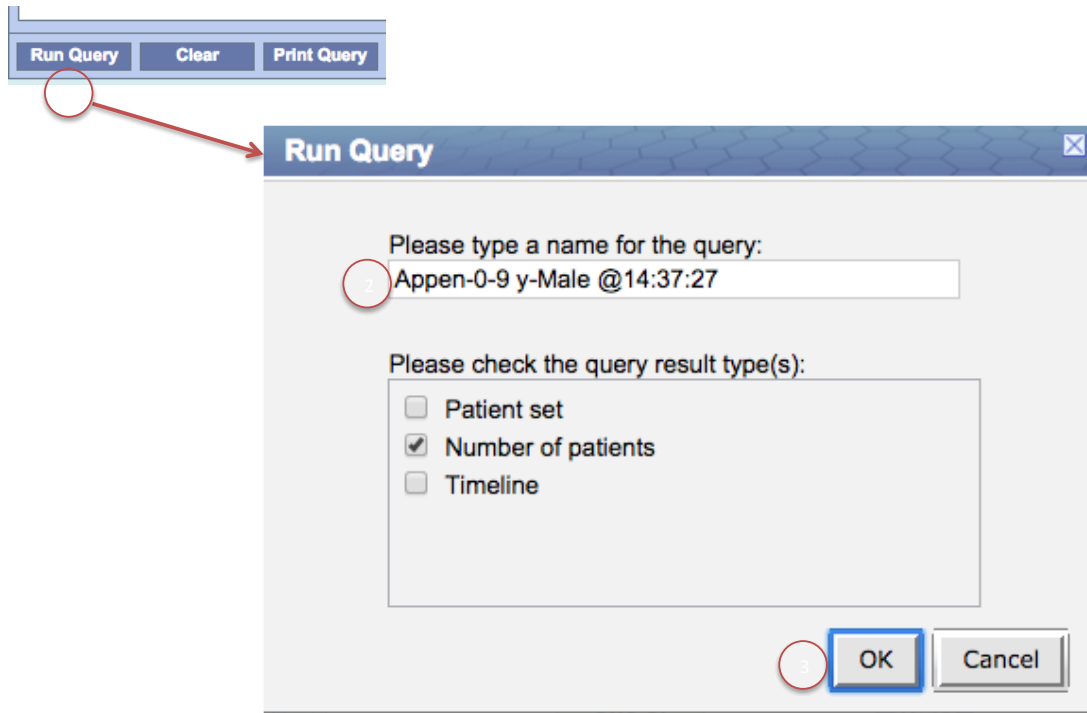
Group 2: (age NOT *0-9 years old* OR *10-17 years old*)

AND

Group 3: *Male* gender

Run the Query

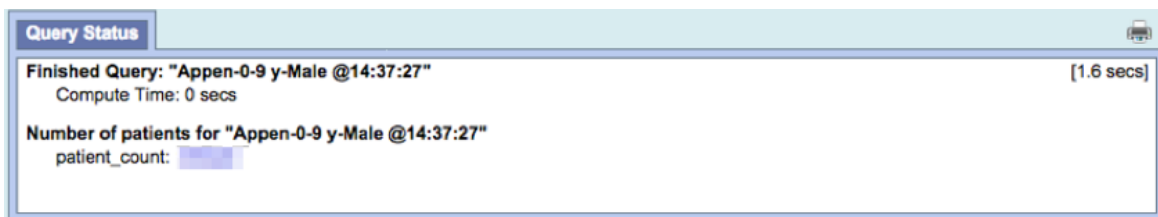
Next, run the query by clicking on the **Run Query** button.



1. Click the **Run Query** button from the **Query Tool** pane.
2. Modify the name of the query for future reference, and make sure that the **Number of patients** box is selected.
3. Click the **OK** button to execute the query.

Review the Query Results

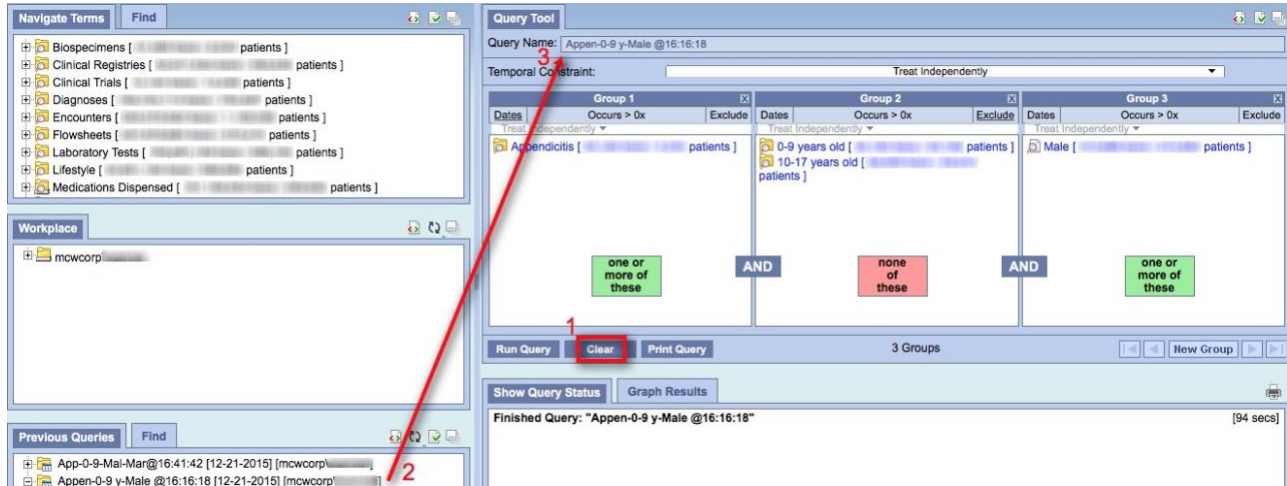
The **Query Status** panel shows the results of the query for the aggregate number of potential patients.



A completed query can become the basis for new queries, for a de-identified retrospective research cohort or for the identification of potential translational research study recruits.

Redefining a Previous Query

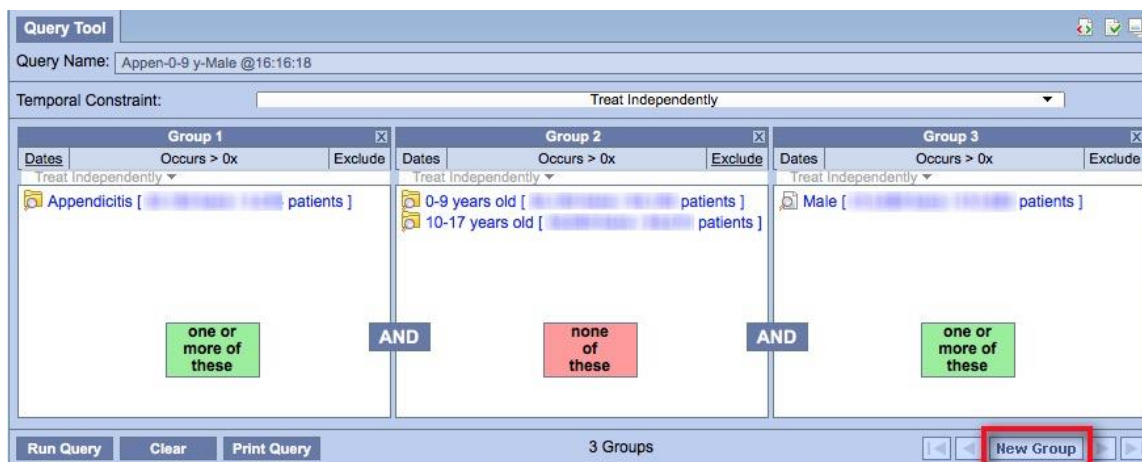
Using our completed query for Adult Male Appendicitis patients as an example, we can use the existing saved query as a starting point so we don't have to re-do previously performed Query Tool steps.



1. First, click the **Clear** button to clear all Group box(es) and start fresh.
2. Click and hold the *Appen-0-9-y-Male* folder in the **Previous Queries** pane.
3. Drag and drop the query folder into the **Query Name** field at the top of the Query Tool pane. A brief **LOADING** status message may be displayed, and then the detailed criteria from the previous query will populate into the Group boxes.

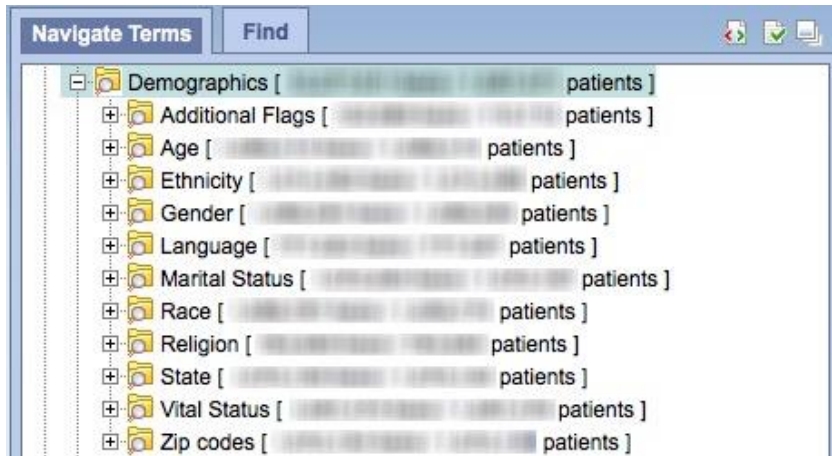
Adding More Query Groups

To add more criteria to an existing query, click the **New Group** button in the lower right corner of the **Query Tool** pane. A new, empty **Group 4** box will appear in the far right box and all existing query groups will be shifted to the left in the **Query Tool** display.

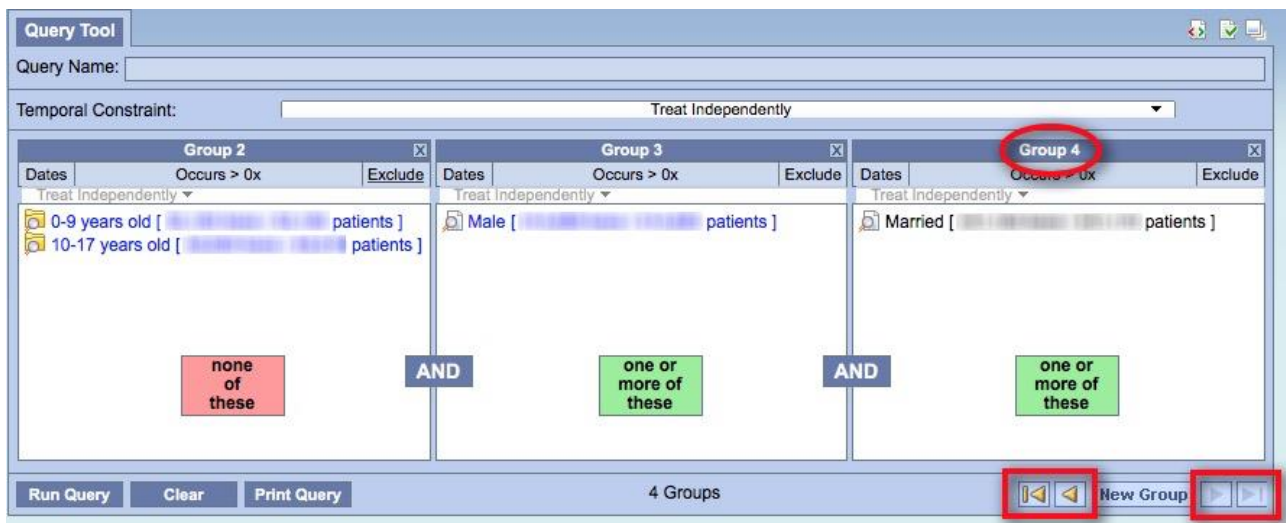


Suppose we are satisfied with the number of patients found in our Adult Male Appendicitis cohort and now want to use this broad cohort as the basis for creating smaller cohorts for comparative analysis.

We can create separate, comparable queries for adult male appendicitis patients with differing demographics such as Marital Status, Ethnicity or Zip Code.

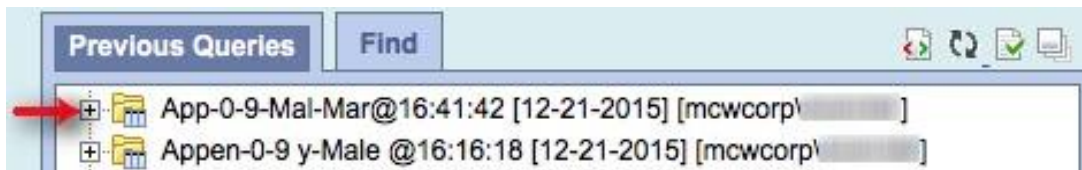


Drag and drop a new criterion from **Navigate Terms** into the **Group 4** box in **Query Tool**.



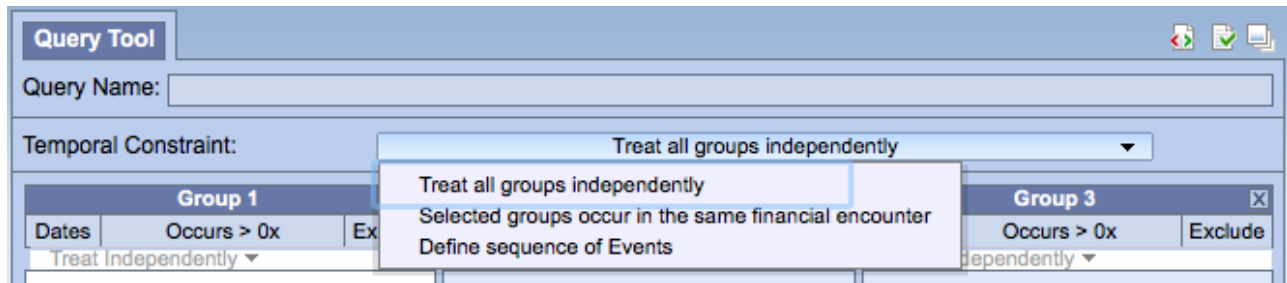
The arrows allow scrolling back and forth through the query Groups as needed.

Click **Run Query** to generate a new, uniquely named cohort query.



Using Temporal Constraints

This function allows users to define time-based relationships between criteria groups or between occurrences within a single group. The default is **Treat Independently**.



Treat all groups independently

This default assumes there are no temporal relationships between criteria groups.

Selected groups occur in the same financial encounter

Do not use this option. Our data warehouse is not yet configured for financial encounters.

Define sequence of Events

This feature allows users to further qualify and join query criteria groups.

Qualifying

Researchers can use the **Events** feature to define a “stable” cohort of patients for retrospective review or study recruitment. Defining and ordering events such as specific diagnoses or encounters sequentially over a meaningful period of time can help exclude patients with inadequate clinical data from data extracts or recruitment lists.

Joining

Researchers can use the **Events** feature to join the occurrences of two criteria groups using the *Occurs Simultaneously* operator.

For Example:

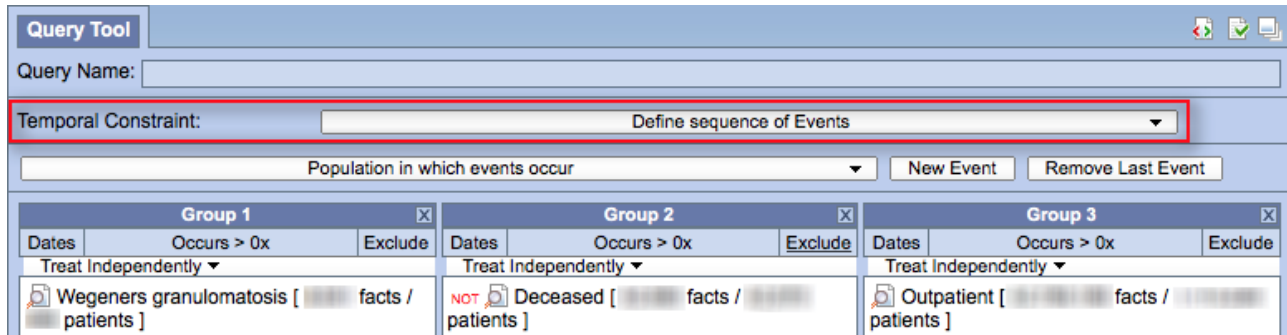
We want to define a stable cohort of patients who have been diagnosed with Wegener’s granulomatosis and have had at least 4 outpatient encounters. We also want to make sure we only find patients who are alive for possible study recruitment.

First, build the cohort query, putting criteria groups in the order of most to least specific:

- Group 1 = Diagnosis of Wegener’s granulomatosis from all sources
- Group 2 = Patients with Vital Status ≠ Deceased (drag Deceased, click Exclude)
- Group 3 = Encounter Type of Outpatient (at least 4 occurrences)

The criteria groups define the **Population in which events occur** for the next steps.

Next, select **Define sequence of Events** from the **Temporal Constraint** drop down list.

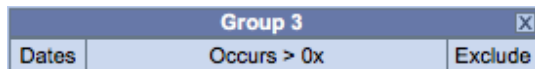


Decide which criteria groups will be defined as an Event and how the Events will be ordered.

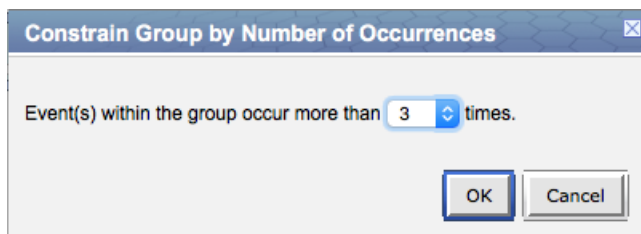
One way to qualify patients as “stable” is to define a diagnosis or type of encounter as an Event and then set time period parameters between certain occurrences of that Event.

For the Wegener’s cohort, stable patients will be defined as those with at least 4 outpatient encounters and more than one year between the first encounter and the last.

In **Group 3** (Outpatient), click the **Occurs > 0x** link in the header to open the selection box.

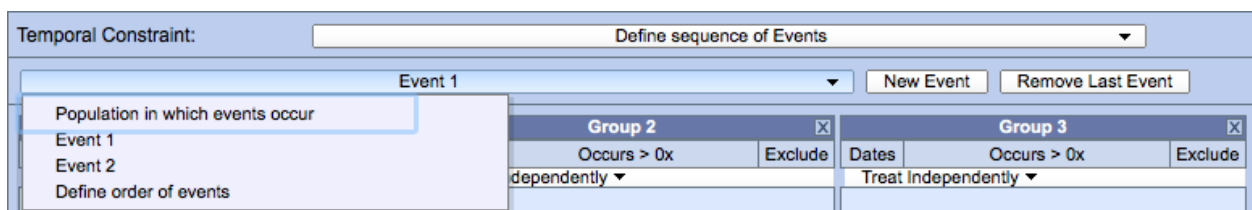


Open the pick list and select 3 to set the number of outpatient encounter events to “more than 3” which equates to “at least 4”



Now we want to define and set the **Event** parameters for “more than one year between the first encounter and the last” using the Outpatient Encounter Type criterion.

Open the drop down list where **Population in which events occur** appears by default. Click **Event 1**. This will transform the **Group 1** box into an **Anchoring Observation** box.



Drag and drop the Outpatient Encounter concept from the Terminology Tree into the **Event 1 Anchoring Observation** box.

Go back to the drop down list where **Event 1** is displayed and select **Event 2**. Drag and drop the Outpatient Encounter concept from the Terminology Tree into the **Event 2 Anchoring Observation** box.

Now that we have defined both **Event 1** and **Event 2** as Outpatient Encounter, we need to set the order of these events so that our cohort only includes patients with an outpatient visit history that spans greater than one year.

Go back to the drop down list where **Event 2** is displayed and select **Define order of events**.

We define the order of Outpatient Encounter events as follows:

The Start of the Last Ever Event 1 [Outpatient Encounter] Occurs After the Start of the First Ever Event 2 [Outpatient Encounter] By greater than 1 year(s).

Note: **New Event** and **Remove Last Event** allow users to insert or delete Event parameters.

Additional event parameter boxes can be inserted or removed using the **Add Temporal Relationship** and **Remove Last Temporal Relationship** buttons.

To leave the Events parameter screen, go back to the drop down list where **Define order of events** is displayed and select **Population in which events occur** to return to the main criteria definition screen.

Note: i2b2 automatically populates the **Define order of events** screen with default values, so be sure to review all parameter fields carefully before running the query.

Click the **Run Query** button, name the query descriptively, and click OK to run the query.

Queries using Temporal Constraints will appear in the Previous Queries tab with a **(t)** prefix before the query name to indicate that temporal criteria have been added.

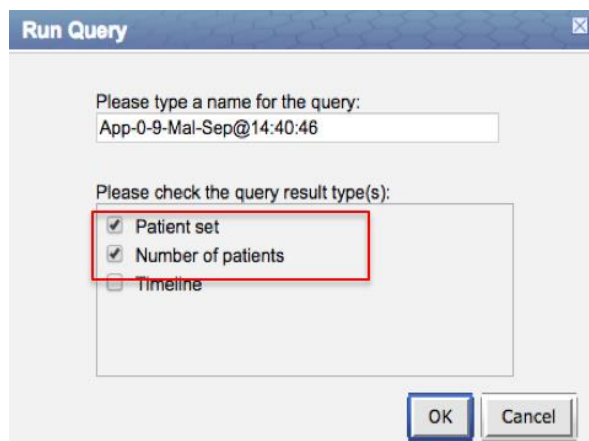
Analysis Tools

Patient Set

The **Patient Set** option within Run Query provides a breakdown of demographic patient information contained within a single query or between two queries.

Consider our cohort: **Adult males diagnosed with appendicitis since 1/1/2005**

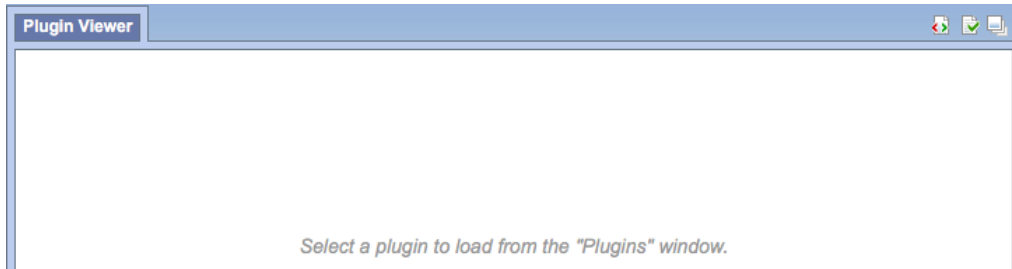
To enable demographics analysis on our cohort, select the query result type options of **Patient set** and **Number of patients** in the Run Query box and click **OK**.



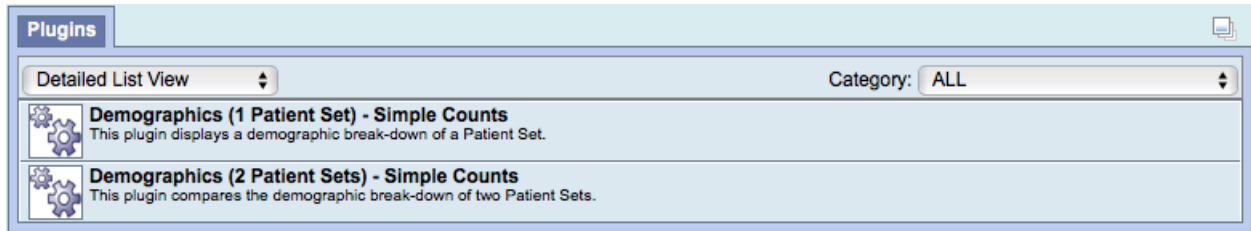
When the query is finished running, click on **Analysis Tools** from the topmost tool bar in i2b2.



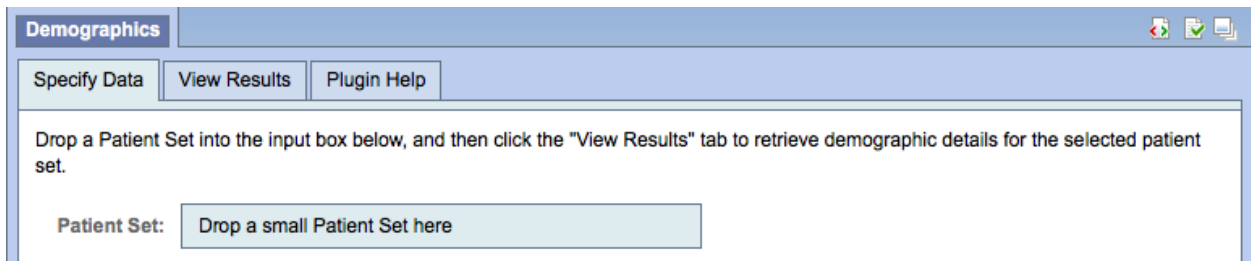
The PlugIn panes will appear and ask you to select a PlugIn from the **PlugIns** pane.



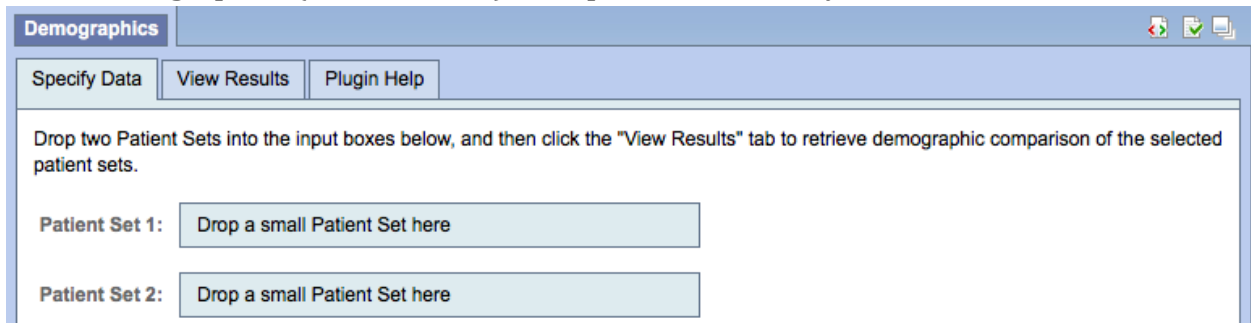
Click on a Demographics box from the **Plugins** pane to activate a Demographics analysis pane.



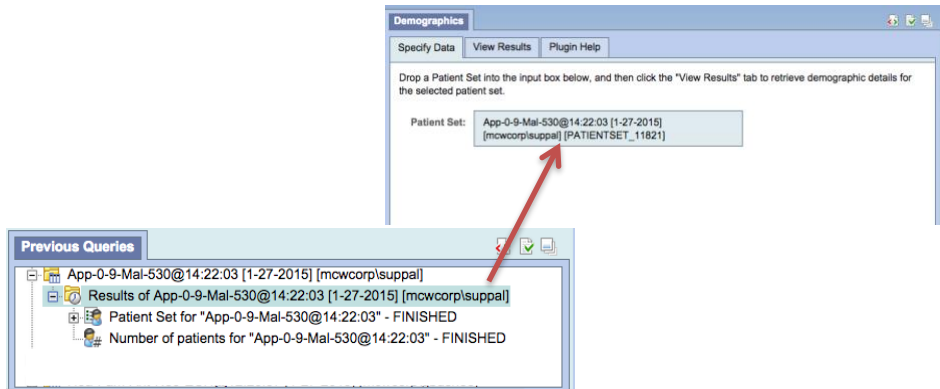
Select **Demographics (1 Patient Set) – Simple Counts** to analyze a single data set.



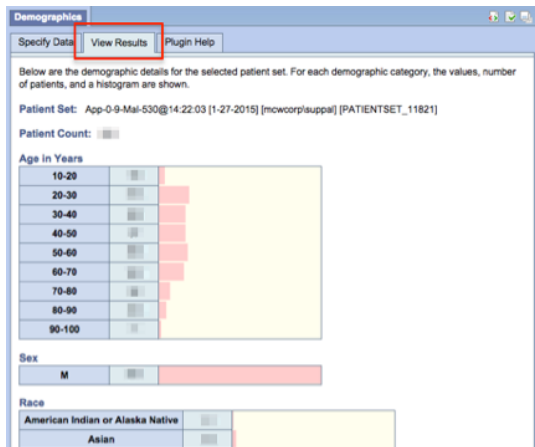
Select **Demographics (2 Patient Sets) – Simple Counts** to analyze data between two data sets.



Drag and drop our Appendicitis query from **Previous Queries** into the box labeled **Patient Set**.



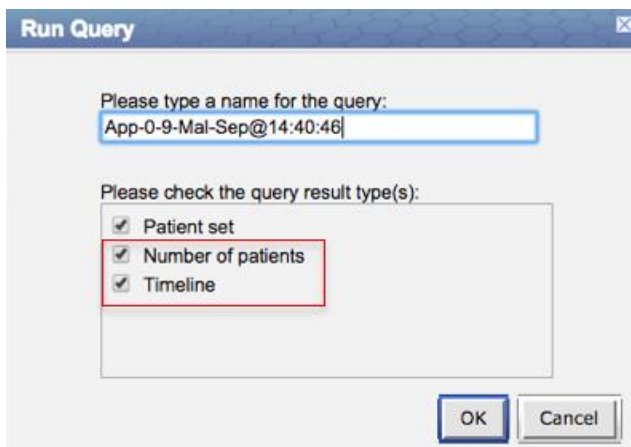
Click on **View Results** to review the **Patient Set** demographics data for the cohort.



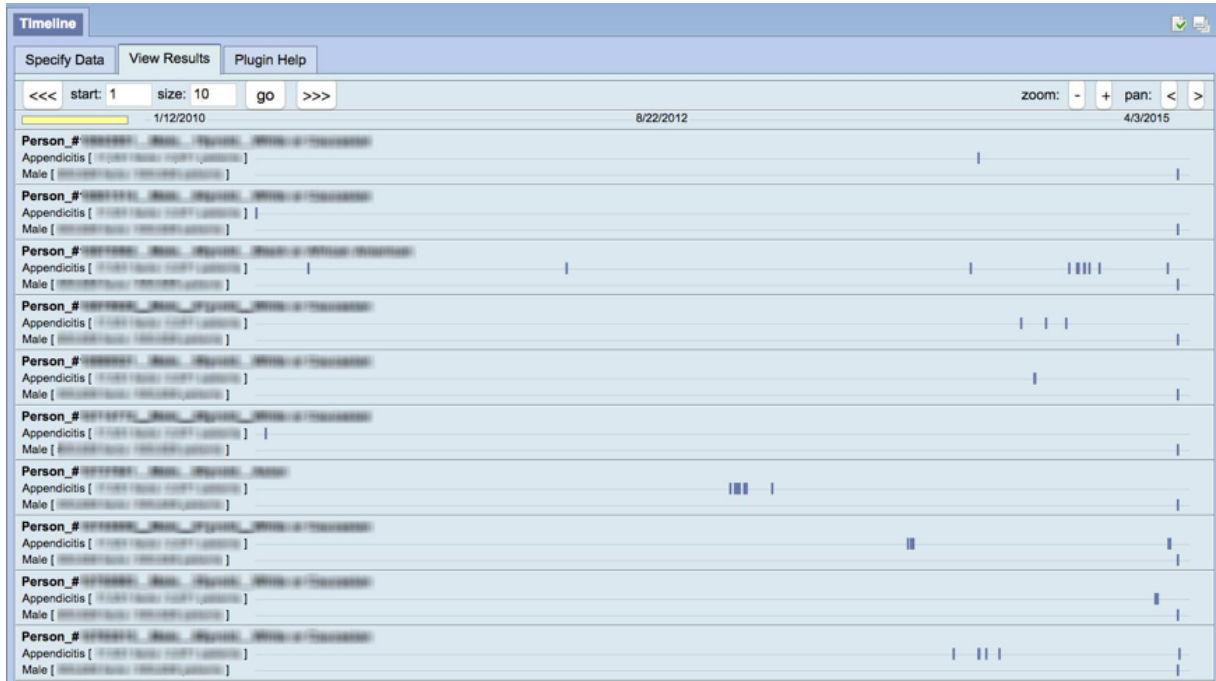
Timeline

The **Timeline** option within **Run Query** provides a graphical time line of recorded events for each patient in a cohort based on the query criteria.

To enable the Timeline graphics for our cohort, select all three query result types **Patient set**, **Number of patients** and **Timeline** in the Run Query box and click **OK**.



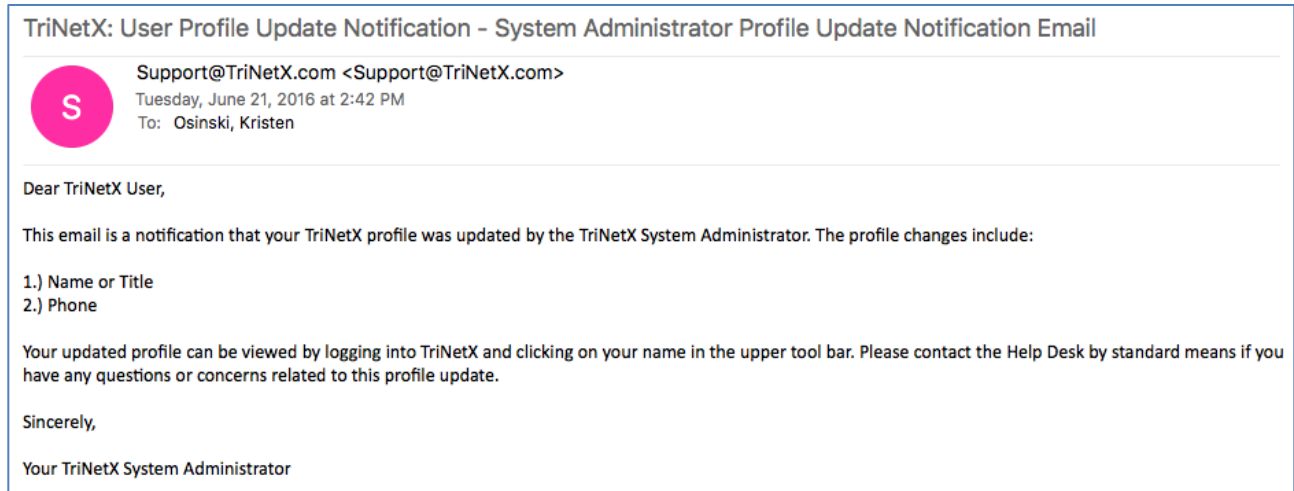
The **Timeline** tab appears automatically and contains a report with data points for each patient in the cohort.



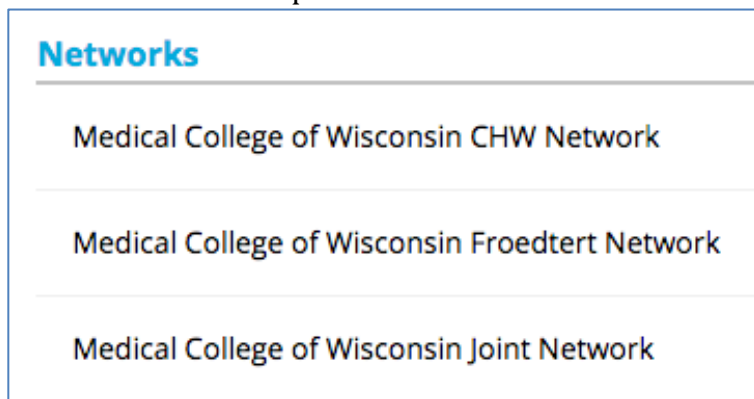
TriNetX

Current i2b2 users will eventually be provisioned with access to TriNetX.

When you have been added as a TriNetX user, you will receive an email similar to the one below.



Users will have access to data from the Children’s Hospital/Specialty Group patient network, the Froedtert Health patient network or from the two networks combined.



Questions and concerns about access to TriNetX should be emailed to the CTSI Biomedical Informatics team (aka CRDW Help Desk) directly at CRDW@mcw.edu.

Questions and issues about the TriNetX tool itself should be emailed to TriNetX Support directly at Support@TriNetX.com

Data Extraction

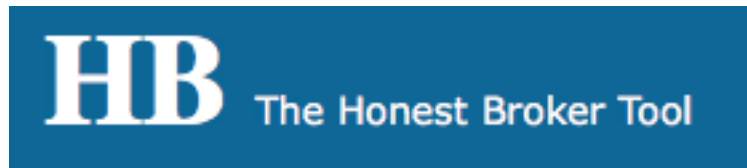
All queries generated in i2b2 are automatically transmitted to the Honest Broker tool.

Researchers can extract data tables from the Clinical Research Data Warehouse in either **identified** or **de-identified** form. The following IRB-driven limitations are enforced to protect our patients from identification:

- Data tables for cohorts of **less than 11** will not be released for de-identified review; however, data tables for these small cohorts can be submitted to the Honest Broker for identified extraction.
- **Identified** data will only be released to researchers with an approved IRB protocol after an IRB delegate reviews and determines that the data requested is appropriate for the protocol.

Data tables for cohorts **greater than 9,999** will not be released for de-identified review to prevent technical performance issues. Contact the CTSI Biomedical Informatics team at CRDW@mcw.edu to discuss and plan for extracting data tables for a very large cohort.

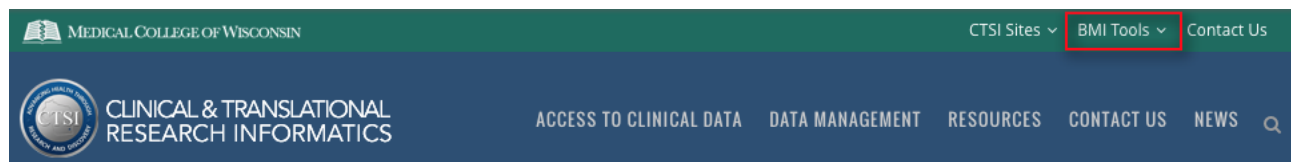
Honest Broker



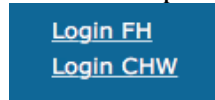
HB access is granted to researchers who complete a [CRDW Data Release Agreement](#). All members of the research team who will have access to data must have current [CITI](#) training.

Log into Honest Broker at <https://i2b2.ctri.mcw.edu/index.html>

HB is also accessible from the ctri.mcw.edu web site under the **BMI Tools** drop down menu.



Note the separate login links for Froedtert Health and Children's Hospital.



Log in with your MCWcorp user name and password.

If you experience problems logging in, email the data warehouse team at CRDW@mcw.edu

Downloading De-Identified Data

Click **Select Patients** to begin the de-identified data extract downloading processes.

[Select Patients](#)

[Select Tables](#)

[HB Extract History](#)

[Extract Agreement](#)

Creating a Query from an MRN or Barcode List

Some research teams know the Epic Medical Record Numbers (MRNs) or MCW Tissue Bank barcode numbers of the patients for whom data is needed for retrospective review or study recruitment. To manually define a query for data extraction:

1. Enter a distinct, meaningful **Query Name**
2. Click the small barcode check box if barcode numbers will be entered as input
3. Enter a list of MRN or Barcode numbers into the **Input** box using one of the formats
4. Click **Select**

Setup a query from MRN/Barcode numbers

1 Please Designate A Query Name *

2 Check if the numbers are barcodes

Please use comma, space, or new line between numbers. (Max number: 9,999)

3 Input the MRN/Barcode Numbers *

4 Select

If all required (*) extract elements are entered, you will be forwarded automatically to the next step: **Select Tables**

Using a Query from Query History

Many research teams use their completed i2b2 queries as the source for data extraction.

Directly select a query from your I2B2 query history

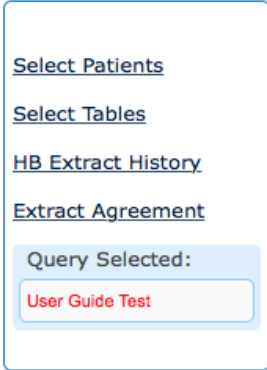
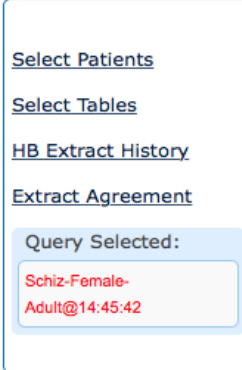
* Please contact us for fetching larger number of patients.
** The patients are less than 10, this query cannot be extracted from de-identification data.

Date	Query Name	Results	Notes
05/24/2016 14:48:04	Schiz-Female-Adult@14:45:42	■■■■	Number of patients for "Schiz-Female-Adult@14:45:42"

Clicking a query hyperlink in the Query Name column will automatically forward you to the next step: **Select Tables**

Selecting Data Extract Parameters

Regardless of which **Select Patients** option is used, the chosen query will appear in red text in the blue **Query Selected** box on the left side of the screen:

Manual Query:  Query from i2b2: 

The manual query screenshot shows a list of options: [Select Patients](#), [Select Tables](#), [HB Extract History](#), and [Extract Agreement](#). Below this is a 'Query Selected:' box containing 'User Guide Test' in red text.

The query from i2b2 screenshot shows the same list of options. Below this is a 'Query Selected:' box containing 'Schiz-Female-Adult@14:45:42' in red text.

First, define the **Extracted Data Date Range** for your de-identified data extract.

Note: if your de-identified extract will be used for subsequent identified data requests, this date range must match the date parameters defined in your approved IRB protocol.

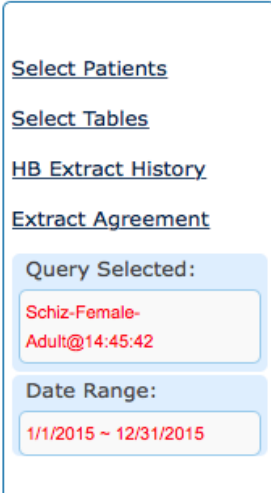
Extracted Data Date Range

Please enter the date range for the data to extract from the Epic tables. (leaving the end date blank, for example, will be all data from the start date to current).

Date Start: ~ End Date:

Set Date

Clicking **Set Date** displays the **Date Range** in red under the Query Selected box:



The screenshot shows the same menu of options as before. Below the 'Query Selected:' box, which still contains 'Schiz-Female-Adult@14:45:42' in red, a new 'Date Range:' box has appeared, containing '1/1/2015 ~ 12/31/2015' in red text.

Warning: if no dates are selected, the extract will contain **all** data available for the cohort! This omission could inadvertently produce very large volumes of data!

Next, select the data tables you want to download for de-identified review.

Use the [+] and [-] icons to open and close the table folders. All data points within a table folder will be downloaded. Some are noted as 'not provided' to prevent patient identification.

Select Tables

Please select the tables you want to extract

- Patient Info
 - Allergy
 - Demographics
 - Family History
 - Immunization
- Encounter Info
 - Patient Encounters
 - Hospital Encounters
 - Reason for Visit
- Lifestyle
 - MRN
 - Associated Date
 - Encounter ID
 - Tobacco
 - Alcohol
 - Illegal Drugs
 - Sexual Behavior
- Laboratory Tests
- Diagnosis
- Procedures
- Medications Ordered
- Medications Dispensed
- Biospecimens

Example:

- Patient Info
 - Allergy
 - Demographics
 - MRN
 - DOB (Age)
 - Gender
 - Name (for deidentification, not provided)
 - Vital Status
 - Address (for deidentification, not provided)
 - Zip Code (for deidentification, only zip 3)
 - Primary Care Provider
 - Marital Status
 - Religion
 - Language
 - Ethnicity
 - Race
 - Family History
 - MRN
 - Contact Date
 - Relation
 - Medical History
 - Encounter ID
 - Immunization

MRN is the common Patient ID data element found in each table and will download as a scrambled character set.

Clicking the box next to a Table folder will display it in red text in the blue **Tables Selected** box and trigger the **Run Extract** button to appear automatically.

[Select Patients](#)

[Select Tables](#)

[HB Extract History](#)

[Extract Agreement](#)

Query Selected:

Schiz-Female-Adult@14:45:42

Date Range:

1/1/2015 ~ 12/31/2015

Tables Selected:

Immunization

Run Extract

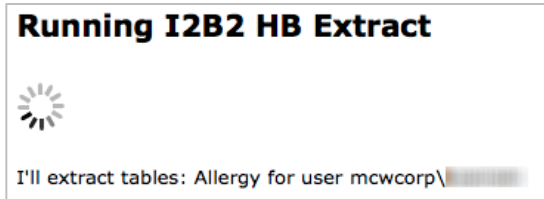
Select Tables

Please select the tables you want to extract

- Patient Info
 - Allergy
 - Demographics
 - Family History
 - Immunization
- Encounter Info
 - Patient Encounters
 - Hospital Encounters
 - Reason for Visit
- Lifestyle
 - Laboratory Tests
 - Diagnosis
 - Procedures
 - Medications Ordered
 - Medications Dispensed
 - Biospecimens

Running a Data Extract

Once all the data extract parameters have been selected, click the **Run Extract** button to generate a de-identified data extract. The following processing message should appear:



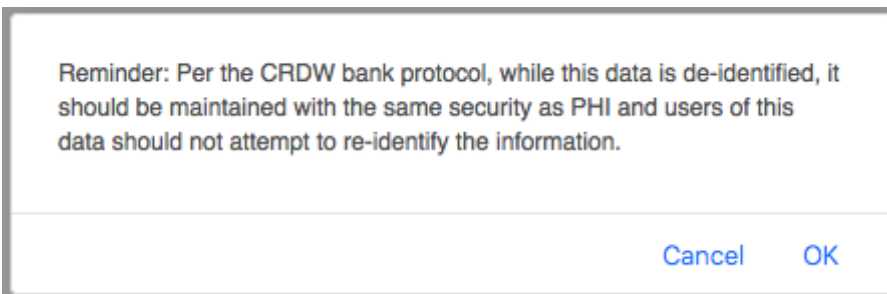
A user-specific **Extract History** table will appear automatically with your most current requests at the top. Extract status will change from **Scheduled** to **Processing** to **Complete**. Processing time will lengthen as the cohort count and/or the number of extract tables increases.

You will receive an email with a hyperlink when the downloading process is complete.

I2B2 De-Identified Extract History							
Note: After one week, extracted files will be removed from the system.							
Queue Date	Query Name	Extract Tables	Count	Status	Finish Date	Link	ID Extract Request
06/09/2016 13:39:30	Fixed Tissue-NephDx@10:17:55	Allergy	19	Scheduled			
06/09/2016 13:30:39	Schiz-Female-Adult@14:45:42	Immunization	2405	Scheduled			
06/06/2016 15:20:08	ILD+PathStatus@11:46:28	Laboratory Tests	15	Completed Reschedule	06/06/2016 15:30:09	ReDownload 58.07 KB	ID HB
06/01/2016 17:33:32	CP Sussex Alive CY2015@16:23:43	Demographics	4487	Completed Reschedule	06/01/2016 17:54:38	ReDownload 217.47 KB	ID HB

If the status for a single table request remains in **Scheduled** status for more than an hour, please contact the CTSI Biomedical Informatics team at CRDW@mcw.edu.

When an extract is complete, a **Download** hyperlink will appear in the **Link** column. Click the hyperlink to download the extracted data to your local workstation. The following pop-up box will appear:



Data output files are delivered in .csv (comma-separated values) format which can be opened and saved in Excel or SPSS. All extracted data files must be saved on a **secure** (encrypted) MCW computer or network drive. **Do not** store this data on portable or flash/thumb drives.

Requesting Identified Data

Click **HB Extract History** to return to your user-specific list of de-identified data extracts.

[Select Patients](#)

[Select Tables](#)

[HB Extract History](#)

[Extract Agreement](#)

Click the **ID HB** button for a previous de-identified extract to request identified data.


I2B2 De-Identified Extract History								
Note: After one week, extracted files will be removed from the system.								
Queue Date	Query Name	Extract Tables	Count	Status	Start Date	Finish Date	Link	ID Extract Request
11/20/2016 10:49:52		Demographics	6	Less 10 Patients				ID HB
11/18/2016 13:56:35	@10:52:38	Diagnosis	407	Completed Reschedule	11/18/2016 14:00:01	11/18/2016 14:01:04	ReDownload 903.53 KB	ID HB


You will be taken to the Identified Honest Broker request screen. When all required fields are complete, click Submit to send the request electronically to the MCW IRB for review and approval.

Identified Honest Broker Request

Please enter the date range for the data to be extracted from the Epic tables. (Leaving the end date blank, for example, will be all data from the start date to current).

Please enter your approved study PRO Number *

Date Start * 

End Date * 

Request Comments

Submit HB Identified Request

If the IRB has questions or concerns about a submission, they will email you directly.

When your submission is approved, you will receive two emails from the Honest Broker: one with a link to the encrypted data table(s) and another with the password to open the .csv file(s).

To access your Identified extract, click the PHI DOWNLOAD link in your My Extracts list.

Store your Identified data on a secure network drive, Office 365 folder or encrypted laptop.

EMERSE Text Search Tool

EMERSE is a document search tool which allows investigators with IRB approval to search for identified text documents from Epic Clarity.

FAQs

What should I do if the i2b2 screen is blank when I log in?

If the i2b2 Concept Tree does not appear within a few seconds after logging in, please email the CTSI Biomedical Informatics team at CRDW@mcw.edu and report that the Concept Tree is blank/empty upon login. Clarify in the email which i2b2 screen is blank: Froedtert (FH) or Children's (CHW).

Can I store de-identified data extracts from Honest Broker on an encrypted flash or thumb drive?

No, all extracted data should be stored on an encrypted MCW laptop or secure network folder.