Foreword: The purpose of this guide is to provide quick pointers to assist getting you started with the CRDB Query Tool (QT). Before starting, please keep the following basic principles in mind:

- The QT is meant solely as a “preparatory-to-research” (PTR) tool to allow casual end-users to define and assess targeted patient cohorts.
- Once cohorts are constructed, users may perform a series of basic queries (or analyses) to gain insights into the types and amounts of data available for the targeted cohort. These basic aggregate counts can be subsequently utilized in research proposals and institutional review board (IRB) applications.

Before starting it is also important to understand what the QT is not. It is not intended to be a tool that can handle every cohort definition. Many query tools have tried to be “all things to all people,” and all have failed. Complex tools require a significant amount of training, which for casual users is all too often forgotten immediately after initial use. The QT’s targeted audience is that of thousands of casual users who will only sporadically need the QT to formulate PTR data. Additionally, the QT does not in itself provide end users direct access to de-identified patient-level data. Use of limited patient-level datasets (LDS) beyond PTR activity requires IRB approval. Projects suitable for LDS data (e.g., de-identified retrospective studies) can often be submitted for expedited IRB review. The QT provides online access to IRB forms on each PRT report output.

Request for complex queries and detailed data are handle through the Office of Informatics and Systems Develop and can be initiated via the “Data Request” tab within the CRDB.
**Getting Started**

**Quick Fact 1: You have to start with something.** One overarching principle of the QT is that of the “base query” or “BQ.” This principle in short means that you have to start with something. In QT processing, your “something” is the BQ functioning as an initial starting patient cohort of interest. A BQ can be narrow (many criteria must be met) or broad (no criteria and start with everyone). BQs are simply a collection of patients and encounters (a cohort) that is passed to every analysis or “reduction” (e.g., filter) process.

The initial QT screen looks like this:

A quick review of the above indicates that all basic parameters are selected. A user could simply enter “All patients” in the query description and press the submit button to create a BQ that encompasses all patients at all encounters. This is perfectly valid and may be appropriate for some queries. However, please know that if you use this strategy, every downstream analysis will include every patient at every encounter, which increases each analysis run time. In general, it will be more efficient and meaningful to downstream processes to provide as much specificity as possible in the BQ. Selecting demographics, sources, date ranges, ICD9 codes and/or diagnosis groups serve to narrow your BQ in terms of patients and encounters. **One final note here: No process can add patients or encounters to a BQ once developed. Reductions can be used to filter or reduce a BQ cohort (or a reduction cohort) to a new cohort – but nothing can augment a BQ cohort.**
Quick Fact 2: **Judicially analyze.** When you submit your initial BQ, the computing cluster performs 8-10 analyses on any subjects in the selected BQ and sends the output (as a PDF) to the e-mail account associated with your user ID. This “output” also can be retrieved using the “O” option associated with the BQ row displayed in your “My Queries” tab.

The “Analyze” or “A” button can be utilized to gather more information on the cohort located by the BQ. There currently are 33 analyses that can be applied to a BQ. Using the BQ row provided in the “My Queries” tab, click on the “A” located to the right of the desired BQ row to select and submit analyses for processing.

**Generate additional analysis of the query results:**

Select modules using the checkboxes located to the left of each desired analysis option. When done, click the submit button at the bottom to return the BQ dataset back to the cluster for additional processing. Once the processing is complete, you will be sent, via e-mail, a new PTR report. The output is also available on the “O” option. A few rules for analyzing BQ:

1. If you intend to perform reductions on CPT or ICD9 (DX) codes, then performing additional analyses (CPT or DX) may help provide data on your intended target codes. **Results from these analyses populate your reduction options to help gate your downstream processes.**

2. Avoid analyses that target “ALL.” These analyses will return results that include the entire population on values in the target value set (e.g., CPT, DX, lab codes, etc.). Output may be MANY pages depending on the target value set.

3. Perform what you need but don’t get carried away. Resist the urge to just select all options. The cluster has a resource governor to prevent a single user from consuming all resources. When you submit more analyses, the overall effect is that your run time is extended. Good cluster etiquette is that you run what you need but don’t hog the resources.
**Quick Fact 3: Reduce a cohort to a more targeted (smaller) cohort.** An intentional design element of the QT (and the entire CRDB) is the notion of starting with a BQ and then “reducing” it to smaller cohorts as information is disclosed about the initial BQ cohort. The overall process is that of a filtering out of patients as reductions are applied. This is an important concept to remember, as there is no operation in the QT that can add patients to a BQ once it is defined and located.

Here is a simple example of using reductions: If one were looking for all male patients over the age of 50 with tachycardia (ICD9s 427.0, 427.1 and 785.0) who had a pacemaker implanted (CPT=33216, 33217,33218) sometime between 1/1/2007 and 12/31/2015, the following strategy could be used:

1. **Step 1.** Define a BQ, using the initial QT page, with all the parameters except those of the CPT. Submit this to the cluster and in about four minutes you get a detailed PTR report with 10 data analyses showing 7,310 patients.

2. **Step 2.** Use the “reduction” or “R” option for the BQ created in Step 1 (found on the “My Queries” tab) to request a reduction by CPT code. You would then use the checkbox options to select CPT codes 33216, 33217 and 33218. After selecting CPTs, submit the form. In a few minutes, you will have a new sub-query on the “My Queries” tab that is indented and located directly below the parent BQ. The reduction query holds the sub-population that has the desired CPT codes. In this case, you would have 86 patients in a new cohort.

This process can be repeated at any level, many times. A single BQ can have multiple reduction cohorts and secondary reduction cohorts can be reduced to third- or fourth-level cohorts. All options (e.g., requests for additional analyses) and functions can work at any level (BQ or reduction). In effect, one can hone in on a population through reductions. When you get to your lowest level of interest, stop reducing.

From a technical stand-point, the computing cluster holds all BQ, and reduction queries as independent collections of patients and encounters. Each cohort dataset is cached and maintained on the cluster and can be referred to through a combination of your user ID and your associated job ID. Since these datasets all are formatted the same and reference the same data (just different collections), they can easily be pushed to analytic routines. Additionally, routines that operate on one dataset will work on all derivative datasets. Informatics personnel also can access these datasets behind the scenes to perform data extractions and sophisticated analyses.
Quick Fact 4: **Know your codes.** Often, it is difficult to develop appropriate queries if you cannot express inclusion criteria with an appropriate set of codes. There are numerous code sets - procedure codes (CPT), diagnosis codes (ICD9/10), diagnosis related group (DRG) codes, lab component identifiers, etc. Clinical informatics professionals often have to perform some analyses to get the correct codes and logic as some codes may be specific to an institution and encounter type. In many cases, clinical departments will have medical coders that assist in the CPT and ICD9/10 coding of clinical encounters, and these individuals can be of great assistance in locating the codes you may need. Selecting CPTs or ICD9 codes without some knowledge of how these are actually utilized in the institution may lead you to incorrect results.