STRIDE Cohort Discovery Tool User Guide

Last updated in January, 2011

TABLE OF CONTENTS

- Overview
- The STRIDE Clinical Data Warehouse
  - STRIDE CDW Statistics
- Launching the Application
- Application Controls
  - The Toolbar
  - The Search Elements
  - Active Search Criteria
    - Filtering By Diagnoses
      - An Example Search
      - Performing AND or OR Searches
  - Filtering by Procedures
  - Filtering Clinical Documents
  - Filtering Lab Orders and Results
  - Filtering with Temporal Constraints
- The Footer
- Protecting Patient Privacy
- Getting Help
- Terms of Use

The STRIDE Cohort Discovery Tool

Overview

To assist SUMC researchers in determining if the STRIDE Clinical Data Warehouse (CDW) contains data suitable for research studies or to help in identifying potential research patient cohorts, we have developed a self-service computer program called the STRIDE Cohort Discovery Tool (CDT). This computer program allows SUMC researchers to directly search the STRIDE CDW using one or more of the following criteria:

- Patient Demographics (Age, Gender, Race, Ethnicity)
- ICD-9-coded diagnoses (inpatient and outpatient)
- ICD-9 and CPT coded clinical procedures (inpatient and outpatient)
- Laboratory test panels/batteries performed and numeric lab test results
- Pharmacy orders
- A variety of clinical documents, including:
  - Clinical notes
  - Procedure reports

RESOURCES AND ASSISTANCE

HELP: Call the IRT Help Desk at 5-8000 or e-mail cci-help-info@lists.stanford.edu for assistance with the STRIDE Cohort Discovery Tool.

CONSULTATIONS: To more about using SUMC data from the STRIDE Clinical Data Warehouse via a free Informatics Consultation - see http://clinicalinformatics.stanford.edu/consultation/
To use the STRIDE Cohort Discovery Tool you must be:

- A SUMC faculty member or Senior Academic Research Staff
- Connected to the SUMC Network (School of Medicine, LPCH or SHC). You must use a **VPN** from remote locations.
- Running Mac OS X 10.4.9 or later, or Windows XP (with SP2), Windows Vista, or Windows 7
- Using a computer with Java JRE version 1.5.0 or later installed. To check your computer, visit [www.javatest.org](http://www.javatest.org).

The output of this tool is an approximate number of patients who meet the entered criteria. It also provides basic demographic statistics on the resulting patient cohort.

**THE CDT DOES NOT EXPOSE ANY DEMOGRAPHIC OR CLINICAL DATA ON INDIVIDUAL PATIENTS.**

Search criteria can be saved and used later as part of the STRIDE Clinical Research Data Review process, or to request clinical data sets from the STRIDE CDW for research purposes (following IRB approval).

The STRIDE Clinical Data Warehouse

The STRIDE Clinical Data Warehouse (CDW) contains clinical and demographic information on patients cared for at Lucile Packard Children’s Hospital (LPCH) and Stanford Hospital and Clinics (SHC). The intent of aggregating this information is to support Stanford University Medical Center’s (SUMC) clinical and translational research mission. Access to patient data in the STRIDE CDW for research purposes requires Stanford IRB approval.

**STRIDE CDW Statistics**

The STRIDE CDW contains the following data through January, 2011, though not necessarily on all patients:

- 1.6 million pediatric and adult patients with clinical and demographic data (1994 - present)
- 14.7 million Clinical Encounters (1994 – present)
- 23 million ICD-9-coded inpatient & outpatient diagnoses (1994 - present)
- 16 million ICD-9 and CPT-coded inpatient and outpatient clinical procedures (1994 - present)
- 2.4 million radiology reports (2005 - present)
- 1.1 million surgical pathology reports (1995 - present)
- 6 million transcribed clinical documents (2005 - present), including:
  - discharge and transfer summaries
  - admission notes
  - progress notes
  - consultation notes
  - clinic visit notes
  - clinical diagnosis and medication summaries
  - dictated letters
  - procedure and operative reports
- 116 million laboratory test numeric results (2000 - present)
- 6.9 million inpatient pharmacy orders (2006 - present)
- 8 thousand dates of death drawn from hospital and Social Security Administration records

**Launching the Application**

To run the STRIDE CDT, navigate to the launch url at

[http://clinicalinformatics.stanford.edu/projects/cohort.html](http://clinicalinformatics.stanford.edu/projects/cohort.html), read the terms of use, and click on the launch
button to begin.

The first time using the application, you may be prompted to allow the program to install and run - please approve any such requests. There may also be a delay while the program is first downloaded and installed to your computer. The software is self-updating - each time it is run, it checks to see if a more recent version is available.

If you computer meets the requirements outlined, you should now see the login window (Figure 1). Please use your SUNet ID and SUNet password to authenticate.

![Login window](image1)

Figure 1. STRIDE Cohort Discovery Tool login window

If you receive a launch error and are working remotely, please verify you have established a [VPN connection](#) to Stanford and try again. If you are still unable to connect or authenticate, please call the IRT Help Desk at **5-8000**.

**Application Controls**

You should now be looking at the STRIDE Cohort Discovery Tool application window (Figure 2) consisting of four main areas: the toolbar across the top, the search elements on the left, your active search criteria on the right, and the footer across the bottom.

![Application window](image2)

Figure 2. STRIDE Cohort Discovery Tool Application Window

**The Toolbar:**

The Toolbar, located at the top of the window, contains just five buttons.
**New Search**  
This button clears all filters to begin a new search from scratch.

**Count Patients**  
This button 'executes' a search based on the criteria currently applied. The button will be disabled if the current search criteria are invalid (e.g. when a required input for a search element is not completed).

**Statistics**  
The statistics button generates a graphical breakdown of common demographic data for the cohort defined by the current search criteria. Data visualized includes gender, current age, race, and location. Click on the 'x' in the upper-right corner of the visualization to return to your search criteria.

**Actions**  
The actions button allows you to save your current search criteria and open previously saved search criteria. This version of the STRIDE CDT now stores all saved criteria in the STRIDE database and not on your local computer.

If you have saved cohort queries on your computer from previous versions of this tool, you may email us the 'query.xml' files and we will help you recover your criteria for use in this and future versions of the tool.

Because the STRIDE CDW is frequently updated with new patient information, it is possible for a saved search to generate different results when run at a later date.

**Help**  
The help button provides links to a number of resources to help you with your cohort search, including a link to the Clinical Informatics consultation service where you can continue to refine your search terms into a complete data set for your clinical research.

---

**The Search Elements:**

The Search Elements allow you to construct a patient cohort by applying a series of filter criteria. The current search elements are broken into three categories: Demographics, Clinical Events, and Temporal Constraints. To add a criteria to the search simply click on it. You may also add the criteria by clicking and dragging it from the left side to where you want to place it. Some criteria can be added multiple times. Multiple search criteria are evaluated with the logical AND operator - meaning only those patients that meet all criteria will be included in the total patient count.

**Demographics**  
This data category contains Current Age, Gender, Race and Ethnicity. These criteria are NOT required to create a patient cohort search and should ONLY be used if you require that the cohort search be constrained by these criteria. Please note that 'Current Age' calculates the patient age at the time of the search whereas 'Age at event' in a Clinical Events filter calculates patient age at the time of the relevant clinical event.

**Clinical Events**  
This data category permits you to search based on a variety of clinically relevant information, including: diagnosis, procedure, lab test performed, numeric lab results, drug ingredients or brand, drug class, contents of clinical documents, and vital status. In addition, each event can be further filtered by specifying the patients' age at the time of the event or the date range in which the event occurred. You can search for the intersection of two events (logical AND) by adding a filter more than one time to your search. You can search on the union of two events (logical OR) by clicking on the 'plus' sign on the right side of an active clinical event filter. Please see the search examples below for more details.

**Temporal Constraints**  
Each clinical event in the STRIDE CDW has a date associated with its entry. Using the new Temporal Constraint filter, you can now refine your search based on the relative order or gap in time between a pair of events. For example, you can now easily find patients who took two different drugs within one month of each other over the entire timespan of the STRIDE CDW.
Active Search Criteria:

To build your cohort, you select filters from the Search Elements area and drag them to the active search box in gray. Alternatively, you can double-click on a search filter and it will be added to your active query. Once a filter is added, there are a number of options to configure and customize. The following is an illustration of one such filter (figure 5):

The following attributes are present in most Filter Elements or can be extrapolated to additional filter types.

A Presence or Absence of the Clinical Event
Each clinical filter allows you to select for both the presence or the absence of an event. In this case, by clicking on section A you can toggle between a 'Diagnosis of (include only patients with the specified diagnosis)' or 'No Diagnosis of (exclude any patients with the specified diagnosis)'.

B Description of Event
Begin by entering in your filtering criteria - many filters use an auto-complete feature to help you complete your query by displaying a list of matching terms. For example, start entering 'sore throat' and the auto-complete will show you the correct ICD-9 billing code of '034 Streptococcal sore throat and scarlet fever'.

C Age at Event
The age at event filter is present on all clinical event filters and can be toggled between four states: any age, less than or equal, greater than or equal, and between two dates (inclusive). This filter allows you to specify a patient age or age range during which the specified clinical event must have occurred (or not occurred). The age can be further refined to years, weeks, or days. For example, setting Age at Event is less than or equal to 18 years with an ICD-9 description of 277.0 Cystic fibrosis would find only those patients who had an ICD-9 diagnosis of Cystic Fibrosis when they were aged 18 years or less. Age at Event is a useful way to distinguish pediatric and adult patient cohorts.

D Event Date
This condition allows you to specify a date or date range that will be used to decide which clinical events (diagnoses, procedures, documents and laboratory test results) are considered when applying the filter. Each diagnosis, procedure, document and laboratory test result received by the STRIDE CDW has a date stamp attached, identifying when that data was created – i.e. the date of that "data event". For example, the query (ICD-9 Diagnosis is 277.0 Cystic fibrosis) AND (Encounter Date is on or after 01/01/2000) would include only those patients with a diagnosis of Cystic Fibrosis on or after January 1 2000.

E Add/Remove Description
It is possible to add additional search terms to your filter, for example to search for '034 Streptococcal sore throat and scarlet fever' OR '462 Acute pharyngitis' in a single Diagnosis filter. These terms will be evaluated with the logical 'OR' clause, so either term will match. If you want to further limit a search by saying a patient must have BOTH '034 Streptococcal sore throat and scarlet fever' AND '462 Acute pharyngitis', you would create two Diagnosis filters - one for each ICD-9 code.

F Information
By holding your mouse icon over the information icon, you will see approximate statistics for the number of records in the STRIDE CDW for this filter.

G Remove Filter
You can remove a filter from your search by clicking on the close 'x' icon in the upper right. This also applies to closing the summary statistics window.

The Footer:

On the left side of the footer is a toggle button to hide or show the Search Elements window. Use this to increase the working size of your active search filters. Your name is displayed in the footer, reminding you that all activity on the STRIDE Cohort Discovery Tool is tracked and audited. Please review the Terms of Use at the bottom of this document for more details. Lastly, a feedback link opens a web form where you can provide us with your comments and suggestions.

Creating a Cohort
A patient cohort is created by first starting with all patient details in the STRIDE CDW and filtering for those with or without a certain property. The CDT provides you with approximate numbers for each filter you apply. When satisfied with the results of your query, you may save the query using the Action button in the toolbar.

Filtering By Diagnoses

The International Classification of Diseases (ICD) is widely used to code patient diagnoses for reimbursement, statistical, administrative and clinical needs. SUMC uses trained coding personnel to review patient charts and abstract ICD-9-CM codes following inpatient care. While neither ICD itself nor the human coding process is perfect, ICD codes are the most widely used system for capturing patient diagnoses in a standardized way. One advantage of ICD coding is that, at least theoretically, all patients that share a diagnosis are coded in the same way. ICD uses a shallow hierarchy of codes to represent a general diagnosis and more specific variants. For example Primary Pancreatic Cancer is represented in ICD as follows:

- Malignant Neoplasm of Pancreas (157)
  - Malignant Neoplasm of Head of Pancreas (157.0)
  - Malignant Neoplasm of Body of Pancreas (157.1)
  - Malignant Neoplasm of Tail of Pancreas (157.2)
  - Malignant Neoplasm of Pancreatic Duct (157.3)
  - Malignant Neoplasm of Islets of Langerhans (157.4)
  - Malignant Neoplasm of Other Unspecified Sites of Pancreas (157.8)
  - Malignant Neoplasm of Pancreas, Part Unspecified (157.9)

Note: by selecting an integer parent code, all child codes are included (e.g., 157 will include 157.0, 157.1, ...)

The STRIDE Cohort Discovery Tool also allows you to search for ICD 'E codes' and 'V-codes' as well. Though not strictly diagnoses, these codes logically fit into the general model of using disease states (including injuries, accidents, poisonings, drug adverse effects, medical and surgical misadventures) to define patient cohorts.

ICD-9-CM E-codes (external causes of injury and poisoning codes) are intended to provide data for injury research and evaluation of injury prevention strategies. E codes capture how the injury or poisoning happened (cause), the intent (unintentional or accidental; or intentional, such as suicide or assault), and the place where the event occurred.

ICD-9-CM provides V-codes to deal with encounters for circumstances other than a disease or injury. The Supplementary Classification of Factors Influencing Health Status and Contact with Health Services (V01.0 - V84.8) is provided to deal with occasions when circumstances other than a disease or injury (codes 001-999) are recorded as a diagnosis or problem.

A good source for ICD-9-CM information is:


An Example Search

Drag a new Diagnosis filter to your active search. Click in the text entry area of the criteria and type either an ICD code (e.g. 157.3) or an ICD term (e.g. Malignant Neoplasm of Pancreatic Duct) into the text field of the condition. As you type, the application will begin searching for matching ICD terms and display them in a popup menu:

A lot is going on here behind the scenes. STRIDE will look at the text (or code) that you enter and attempt to interpret it, producing a list of suggested ICD codes that you can choose from. It supports the use of synonyms, so that, for example, entering breast cancer finds "breast neoplasms". Word order and case
are ignored, so that "breast cancer" and "CANCER BREAST" are equivalent. STRIDE will also attempt to display the suggested ICD codes with the most general code at the top of the list.

Selecting a general ICD code (e.g. Malignant neoplasm of the pancreas) instructs STRIDE to include patients whose disease was coded with that ICD code or ANY of its more specific (children) codes. To be precise, selecting ICD code 157 Malignant Neoplasm of the Pancreas will instruct STRIDE to also search for 157.0, 157.1, 157.2, 157.3, 157.4 etc. Unless you are sure that you only want to include patients with a very specific diagnosis (e.g. 157.3 Malignant neoplasm of the pancreatic duct), it is often a good idea to select the more general (or parent) ICD code, as this will find all child codes for the disease.

Performing AND or OR Searches

The plus icon (E in figure 6) allows you to add another search term to a query. In the example above, if you wanted to search for patients with either of two specific ICD-9 codes, 157.0 OR 157.3, you would add a second search term to you query:

Union Queries - searching for A OR B

You can also use the OR search feature to include patients who have at least one of a number of diagnoses. For example if you are interested in patients with smoking-related cancers, one might create the following condition:

Multiple OR clauses to find smoking-related cancer patients.

On the other hand, if you wanted to find only patients with both of these diagnoses, you would add the diagnosis filter two times to the query window as illustrated below:

Intersection Queries - searching for both A AND B

Filtering By Procedures

Medical and surgical procedures performed on patients are coded using ICD and/or CPT (Current Procedural Terminology – see below) codes. In general inpatient procedures performed at SUMC are coded using ICD, while many outpatient procedures are coded using CPT. The STRIDE Cohort Discovery Tool supports integrated searching of both inpatient and outpatient ICD and CPT coded procedures, using the Procedure condition. This operates in much the same way as the Diagnosis condition, supporting 'search as you type', intelligent lookup of ICD and CPT codes, ability to OR procedure codes as well as AND procedure codes. A major difference from diagnosis is that procedure code lookups may return a mixture of ICD and CPT procedure codes. The codes returned are displayed in three sections: CPT Hierarchical Codes ("CH" codes); Individual CPT Codes and ICD-9 Procedure Codes.
Procedural lookup illustrating mixture of ICD and CPT codes

When using the procedure condition in a cohort query you may wish to OR equivalent ICD and CPT codes together in a compound condition to ensure that you include patients who had a procedure performed as either an inpatient or outpatient e.g.:

CPT (Current Procedural Terminology) codes are categorized into three groups:

- **Category I** CPT codes describe a procedure or service identified with a five-digit CPT code (e.g. 29870) and descriptor nomenclature (Arthroscopy, knee, diagnostic, with or without synovial biopsy). The inclusion of a descriptor and its associated specific five-digit identifying code number in this category of CPT codes is generally based upon the procedure being consistent with contemporary medical practice and being performed by many physicians in clinical practice in multiple locations.

- **Category II** CPT codes are intended to facilitate data collection by coding certain services and/or test results that are agreed upon as contributing to positive health outcomes and quality patient care. This category of CPT codes is a set of optional tracking codes for performance measurement. These codes may be services that are typically included in an Evaluation and Management (E/M) service or other component part of a service and are not appropriate for Category I CPT codes.

- **Category III** CPT codes contain a temporary set of tracking codes for new and emerging technologies. Category III CPT codes are intended to facilitate data collection on and assessment of new services and procedures. These codes are intended to be used for data collection purposes to substantiate widespread usage or in the FDA approval process.

**Filtering Clinical Documents**

The STRIDE Cohort Discovery Tool allows searching inside clinical documents for words or phrases, as part of a patient cohort query. A large variety of clinical documents are supported:
Many different types of clinical documents can be searched.

Multi-word text phrases can be searched using three methods:

- **near each other** - this generally means in the same sentence, independent of word order.
- **in the same document** - the words could be anywhere in the document
- **exact phrase** - this means the words must occur together and in order

In general you should choose the default 'near each other' option rather than the 'in same document' option. Case and punctuation are ignored when searching inside documents and word order is ignored for the 'near each other' and 'in same document' searches. If the 'near each other' operator is used, words do not need to be immediately adjacent to match.

**Context and caveats when searching clinical documents**

The utility of searching within clinical documents is somewhat limited by the absence of contextual information. Documents containing the phrase 'Myocardial Infarction' could use this term in many different contexts (e.g. 'a history of myocardial infarction', 'Father died of myocardial infarction', 'Rule-out myocardial infarction', 'Patient has no history of myocardial infarction' etc.). In addition this search would not include the document that states 'The patient says that he had a heart attack two months ago'. In addition, be aware negated terms, e.g. 'no evidence of myocardial infarction' would also be considered a match. We are working on strategies to address some of these issues.

**Filtering Lab Orders and Results**

The STRIDE Cohort Discovery Tool supports using laboratory tests as conditions in cohort searches. Currently the application allows you to use laboratory tests by the type of test performed or by the quantitative results.

**Filter by Test or Battery**

You can including or exclude patients in a cohort based on the presence of a either a laboratory battery or an individual test measurement.

The following search selects patients who had a Hepatic Function Panel (includes Albumin, Alkaline Phosphatase, ALT (SGPT), AST (SGOT), Conjugated Bilirubin, and Unconjugated Bilirubin) or had a single AST test.
Searching for lab batteries or individual tests

The Stanford Clinical Laboratory (www.stanfordlab.com) maintains lists of lab codes and batteries that you may want to search. Many labs can be ordered with a variety of codes so it is common to use the 'OR' functionality to include all common lab variants.

**Filter by Numeric Lab Result**

You can include patients in a cohort only if they have a particular laboratory test numeric result value, or set of numeric result values.

To include a numeric test result in a patient cohort query, use the 'Numeric Lab Test Results' condition. Enter all or part of the name of the test that you are interested in and select a test from the list. Once the laboratory test is selected, you can choose from a menu of operators (see below) as to how you wish to specify the test result.

- Filtering based on lab results
  - You can specify an actual value, a range of values or use 'high', 'low' or 'normal'. You can 'OR' or 'AND' multiple laboratory test results.

**Filtering with Temporal Constraints**

The newly added temporal constraint filter allows for filtering your results based on the proximity or order of events defined in two previous search filters.
Before adding this filter to your query, you should first create two or more clinical search filters that you wish to refine with a temporal constraint. Next, drag or double click on the temporal constraint to add it to your query.

- The two events you wish to compare should be set as Event A and Event B.
- Since each event could actually have many results within it, the ‘event date’ filters allow you to select from three options: Any, Earliest, or Most Recent.
- The ‘order selector’ sets the relative order of the two events and has three options: follows, precedes, or precedes or follows. The last option should be used when relative order does not matter and you are only interested in proximity of events. Swapping events A and B is the same as switching the order selector.
- The ‘range for comparison’ sets the time between the two events and has three options: less than, greater than, or between two values.

For example, if you wanted to determine a Cohort of patients who were prescribed Tylenol within 90 days after a liver transplant, you might formulate the query as:

```
Drug with ingredients including Acetaminophen
  age at event is any age
  event date is any date

Underwent procedure 50.5 Liver transplant
  age at event is any age
  event date is any date

Any instance of Drug ingredients include Acetaminophen
  by between 1 and 90 days
  the most recent instance of Procedure 50.5

~ 290 patients
```

Each search filter may include many events (using the drug Tylenol, for example), it is important to take note of the (Any/Earliest/Most Recent) selector before each entry in the temporal constraint filter. Using 'Any' for both events could lead to confusing results. To search for events occurring on the same day, set the 'order selector' to precedes or follows and the 'range for comparison' to less than 1 day.

**Protecting Patient Privacy**

In addition to never revealing individual patient identifiers or data, the STRIDE Cohort Discovery Tool uses a number of strategies to prevent "triangulation" of data that might identify an individual patient. As a consequence, total cohort sizes of less than ten patients are reported as "<10 Patients" and individual categories in criteria results and demographics graphs are reported in increments of 5, with a small random "fuzzy" rounding factor added to each search results to further prevent triangulation.

**Getting Help**

To get assistance with the STRIDE Cohort Discovery Tool you can call the IRT Help Desk at 5-8000 or send an e-mail to cci-help-info@lists.stanford.edu. You can also view an on-line version of this document by clicking on the help icon in the top right corner of the application window.
Terms of Use

By using the Anonymous Patient Cohort Discovery Tool, participants acknowledge that:

- This system was not designed, nor is it intended, to support any aspect of patient care and its use for any patient care purposes is prohibited.
- The results returned by this system are for SUMC research purposes only and should not be distributed outside of Stanford University Medical Center.
- All searches executed within the system are recorded and may be examined, as part of system evaluation and audit purposes. The identity of the user is recorded along with information related to each search executed.
- No user of the system will share his/her login information with any other person, for purposes of assessing this system.
- All users must be current in HIPAA training and be affiliated with SUMC in order to use the system.