

CARE-conceptDemographics v1.2 User Manual

Introduction

This i2b2 Webclient plugin enables the user to generate a set of demographic histograms on the subset of patients observed with selected concepts (ontology terms).

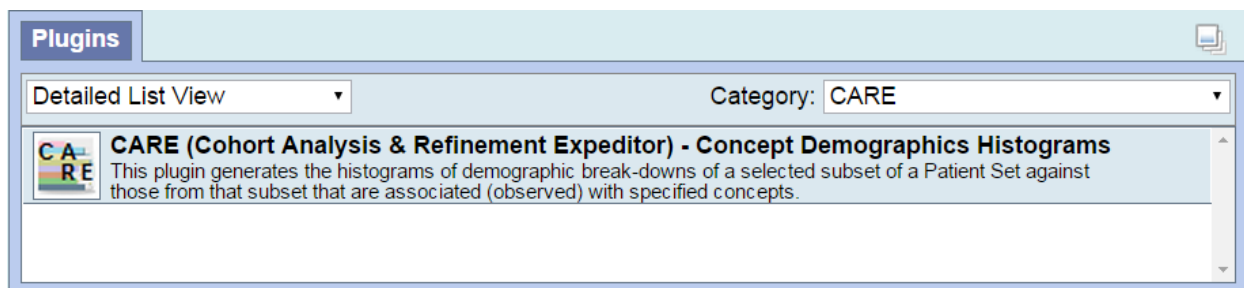
The following are some of the other key functions in this version of this plugin:

- Generates demographic histograms on the subsets of patients associated (observed) with any selected concepts (ontology terms).
- Allows the usage of smaller queries of patient subgroups (from a large patient set) iteratively, in place of one single large query of that huge patient set, to reduce risk of overwhelming the server (i2b2 hive).
- Allows selection of subset of patients from a large patient set, to further reduce risk of overwhelming the server.
- Provides feedbacks on current progress, elapsed time, and estimated remaining run time.
- Provides notification & suggestion in cases of paging (query-subgrouping) by the server.
- Provides termination indication & suggestion (for recovery) when the server failed.
- Provides detailed '**HELP**' and '**HINT**' dialogs in response to corresponding hotlinks strategically placed next to input fields that may need explanations or suggestion.
- Provides options to export histogram data into a CSV or XLS file that can easily be used by MS Excel, SPSS and other statistics software.

This document outlines the usage of this plugin. Please refer to its complimentary document, the “*CARE-conceptDemographics v1.2 Installation Guide*”, for installation details.

Instructions

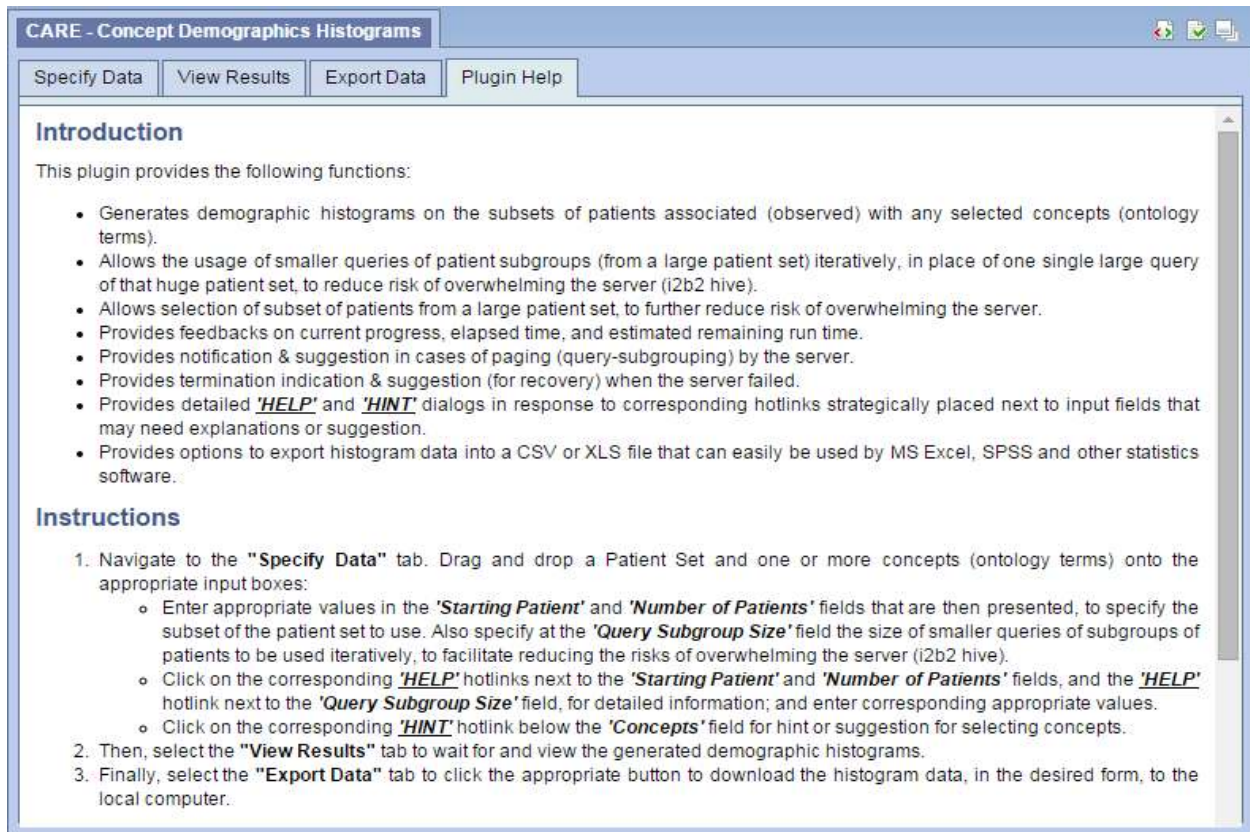
As with any i2b2 Webclient plugins, this plugin is listed under the “Plugins” panel under the “Analysis Tools” tab.



Note that selecting “CARE” from the “Category” pull-down selection list (upper right corner of panel) would bring you to this plugin link faster.

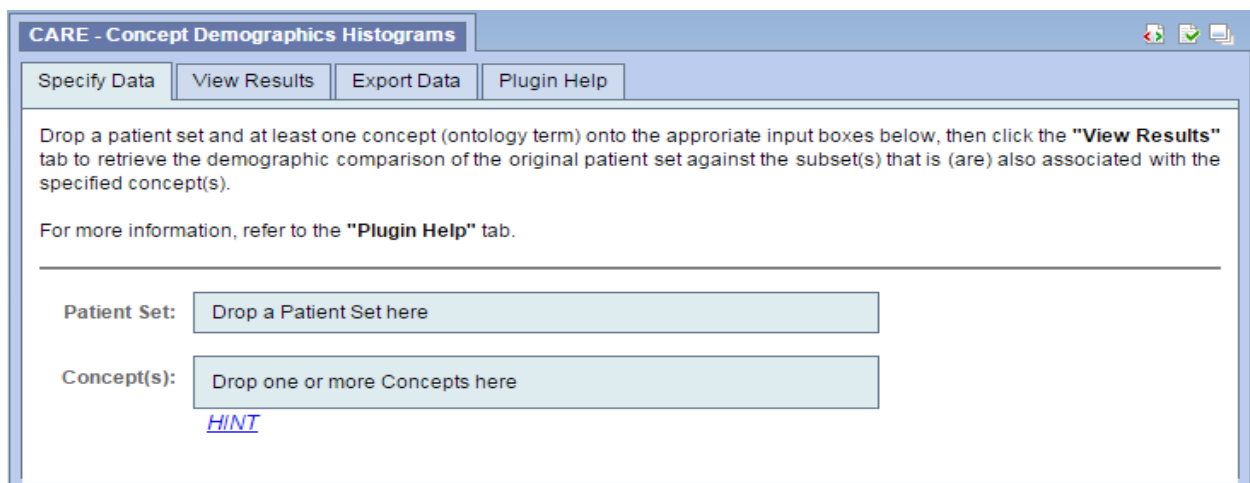
The “Plugin Help” Tab

The first thing a user should do, if this document is not readily available, is to click its “Plugin Help” tab, which contains a summary of the contents in this document.



The “Specify Data” Tab

Navigate to the "Specify Data" tab. Drag and drop a Patient Set and one or more concepts (ontology terms) onto the appropriate input boxes:



As soon as a “Patient Set” is specified (i.e. dropped in), then more fields will be displayed for options input:

CARE - Concept Demographics Histograms

Specify Data View Results Export Data Plugin Help

Drop a patient set and at least one concept (ontology term) onto the appropriate input boxes below, then click the "View Results" tab to retrieve the demographic comparison of the original patient set against the subset(s) that is (are) also associated with the specified concept(s).

For more information, refer to the "Plugin Help" tab.

Patient Set: Diagnoses@15:36:54 [5-31-2012] [demo] [PATIENTSET_65]
contains some 133 patients (using patients #1 - 133, or the entire set)

Starting Patient: 1 Number of Patients: 133 [HELP](#)
Query Subgroup Size: 20 [HELP](#)

Concept(s): Drop one or more Concepts here
[HINT](#)

Select the appropriate 'Starting Patient' (defaulted to 1), 'Number of Patients' (defaulted to 500 or the size of the patient set, whichever is less), and 'Query Subgroup Size' (defaulted to 20) fields that are then presented, to specify the subset of the patient set to use, and the smaller queries of subgroups of patients to be used iteratively, to facilitate reducing the risks of overwhelming the server (i2b2 hive).

Note that when you changed the 'Starting Patient' field value, the ‘using patients # ...’ statement will be updated once the focus is moved to any other field (i.e. if another field is then selected), and the value of the 'Number of Patients' field will be updated accordingly.

Patient Set: Diagnoses@15:36:54 [5-31-2012] [demo] [PATIENTSET_65]
contains some 133 patients (using patients #10 - 133, or just 124 of the set)

Starting Patient: 10 Number of Patients: 124 [HELP](#)

Click on the corresponding '[HELP](#)' hotlink next to the 'Starting Patient' and 'Number of Patients' fields, and the '[HELP](#)' hotlink next to the 'Query Subgroup Size' field, for detailed information (in corresponding pop-up dialogs); and enter corresponding appropriate values.

About 'Starting Patient' and 'Number of Patients':

If your patient set contains thousands of patients, you may want to use just a subset of that. By so doing, you will speed up your query as well as reduce the likelihood of overwhelming the server (i2b2 hive) to the point of failure (and no result).

Also, if you already encountered a server failure (overwhelmed by the combination of large number of patients times lots of concepts), then you may like to rerun this plugin several times. each time specifying different 'Starting Patient' and 'Number of Patients'.

For instance, use 1 and 500, respectively, in your first run; follow with 501 and 500, respectively, in your second run; then 1001 and 500, respectively, in your third run; and so on, until you get enough data.

The example above should also show that the 'Starting Patient' and 'Number of Patients' refer to the patient entry order in the patient set only, which has no bearing on the actual patient IDs.

In addition, if the sum of the 'Starting Patient' and the 'Number of Patients' values exceeds the total patient count, then it'll be adjusted to the remaining patient count accordingly.

OK

About 'Query Subgroup Size':

Please note that a large query (i.e. a large patient set with many concepts, especially ones consisting many subfolder trees) may overwhelm the server (i2b2 hive), which may fail (where no data can be rendered at all) after considerable delays and timeouts. To avoid this problem, you may want to set the 'Query Subgroup Size' value.

Setting the 'Query Subgroup Size' instructs this plugin to temporarily divide up your patient set into subgroups of your specified size, and then iteratively make request of relevant data for each of these resulting subgroups, and then collate and render the returned data (when all these smaller queries are completed).

Since each subgroup should be of small enough size, the requests hopefully would not cause the server to fail or hang.

The ideal 'Query Subgroup Size' value cannot be predicted in general, and strongly depends on the number of observations (related to the total numbers of concepts and their complexities) returned; but values of 20 to 50 may be good. A higher value may result in faster processing but may also carry higher risk of the server failing (where no data can be rendered at all).

Incidentally, setting a value of 0 instructs this plugin not to divide up the original query into smaller queries (and carries the risk of overwhelming and failing the server, as well as practically no 'Elapsed time' and 'estimated remaining run time' updates).

OK

Note that the sample pop-up dialogs are from a Firefox browser, different browsers may have different style dialogs (that are similar but contain the same key contents).

The following pop-up dialog will be displayed when the '[HINT](#)' hotlink, below the "Concept(s)" input box, is clicked, to provide suggestions regarding concept specifications that may result in faster turn-around time:

'Concept' hint (suggestion):

For best results, select finer-grained concepts that may not be related to multiple non-exclusive observations in patients.

For example, while a patient set may be for 'Circulatory system', selecting several finer concepts like 'Hypertensive disease', 'Ischemic heart disease', and 'arterial vascular disease', etc., that are pertinent to the interest at hand, would result in more meaningful results than simply specifying a single concept of 'Circulatory system', which contains other concepts that may be irrelevant to the present study.

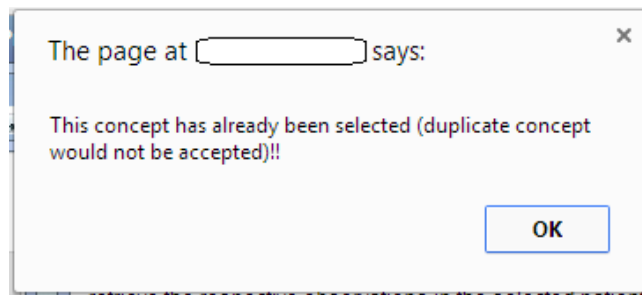
Furthermore, keep in mind that far-reaching concepts (that contains many branches and sub-branches of concepts), coupled with a large patient set, may overwhelm the server (i2b2 hive) to the point of failure, resulting in no data getting returned at all.

OK

As implied in the pop-up dialog above, one can drop in many concepts:

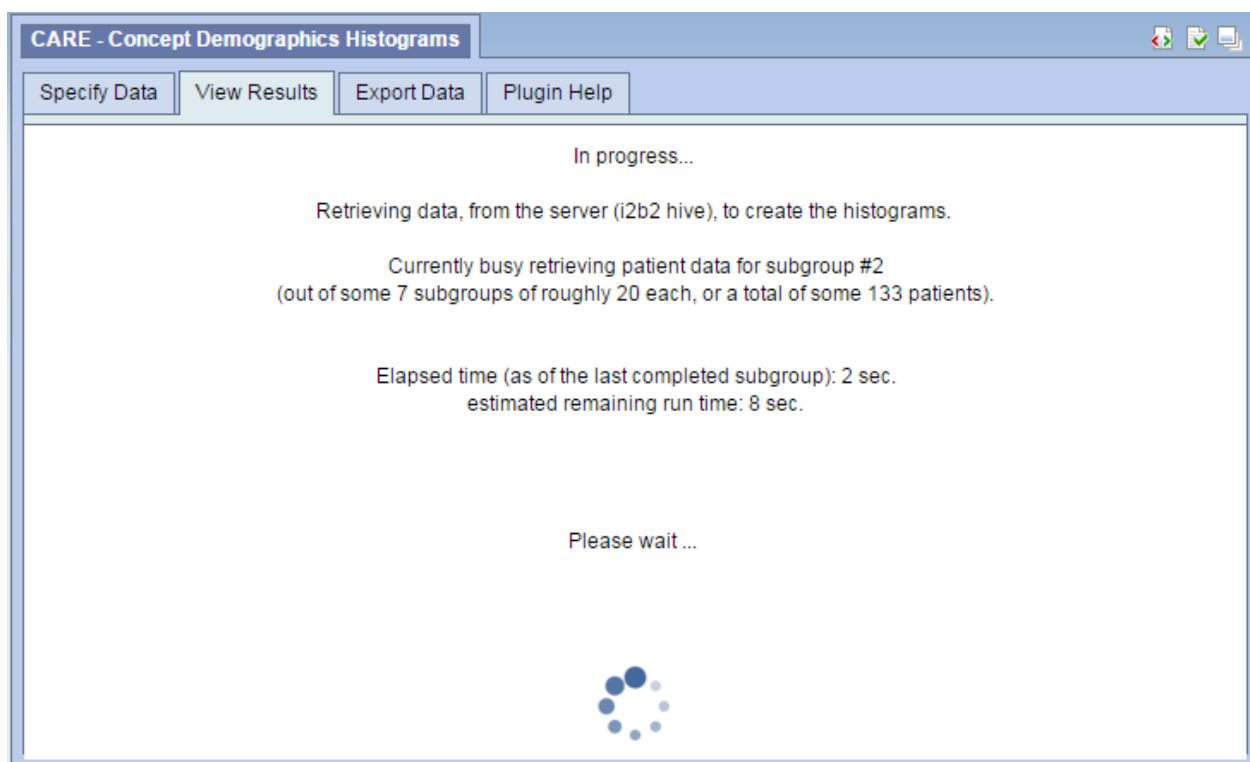
Notice that the instruction "Click a Concept to remove it from the list." is displayed as soon as a concept is dropped into the box.

A warning pop-up dialog will be displayed if a repeated concept is dropped in, similar to the following (from a Chrome browser that was used; note: IP address blotched out):



The “View results” Tab

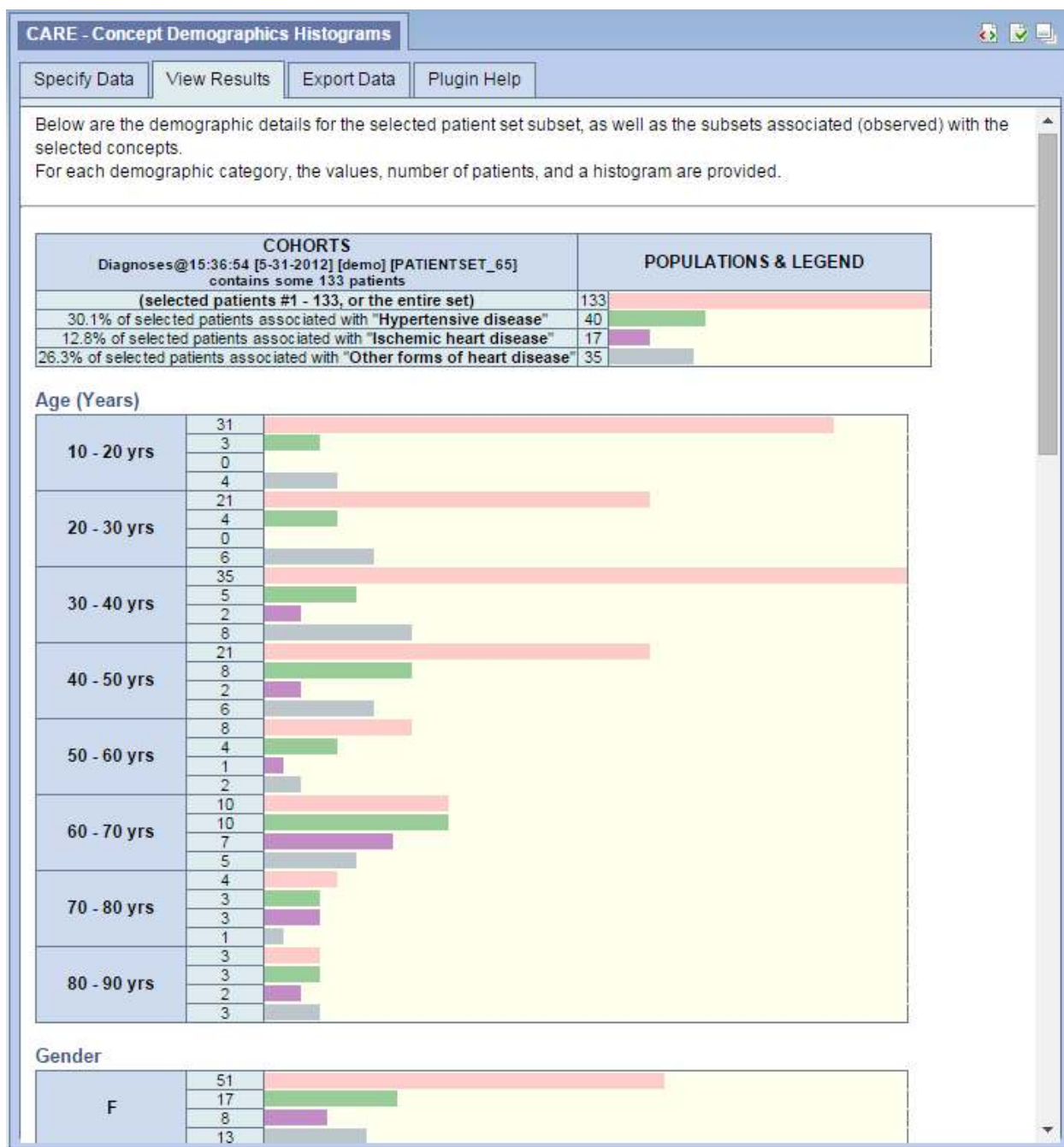
Next, select the "View Results" tab to view the progress and the resulting table of the observations.



Depending on the size of the Patient Set and the number and complexities of the concepts, it could take quite a while for the i2b2 hive to complete the chore requested. However, while the data are being fetched, there will be constantly updated feedback displays to provide estimations as to how much data have been retrieved, how much are yet to be fetched, elapsed time, and remaining run time, etc.

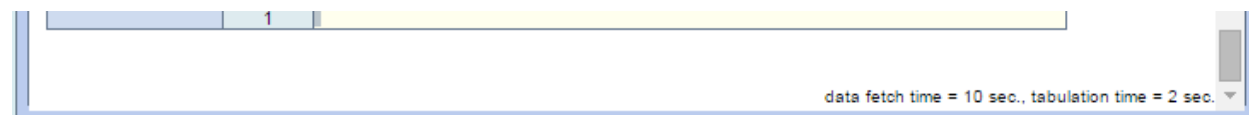
Successful Run

At the end of a successful run, after all the requested data have been returned by the server (i2b2 hive), the resulting histograms will be displayed:



Note that for the figure above, the other histograms not captured in the screenshot (due to clipping) are 'Race', 'Language', 'Marital Status', 'Religion', and 'Vital Status (Deceased)'.

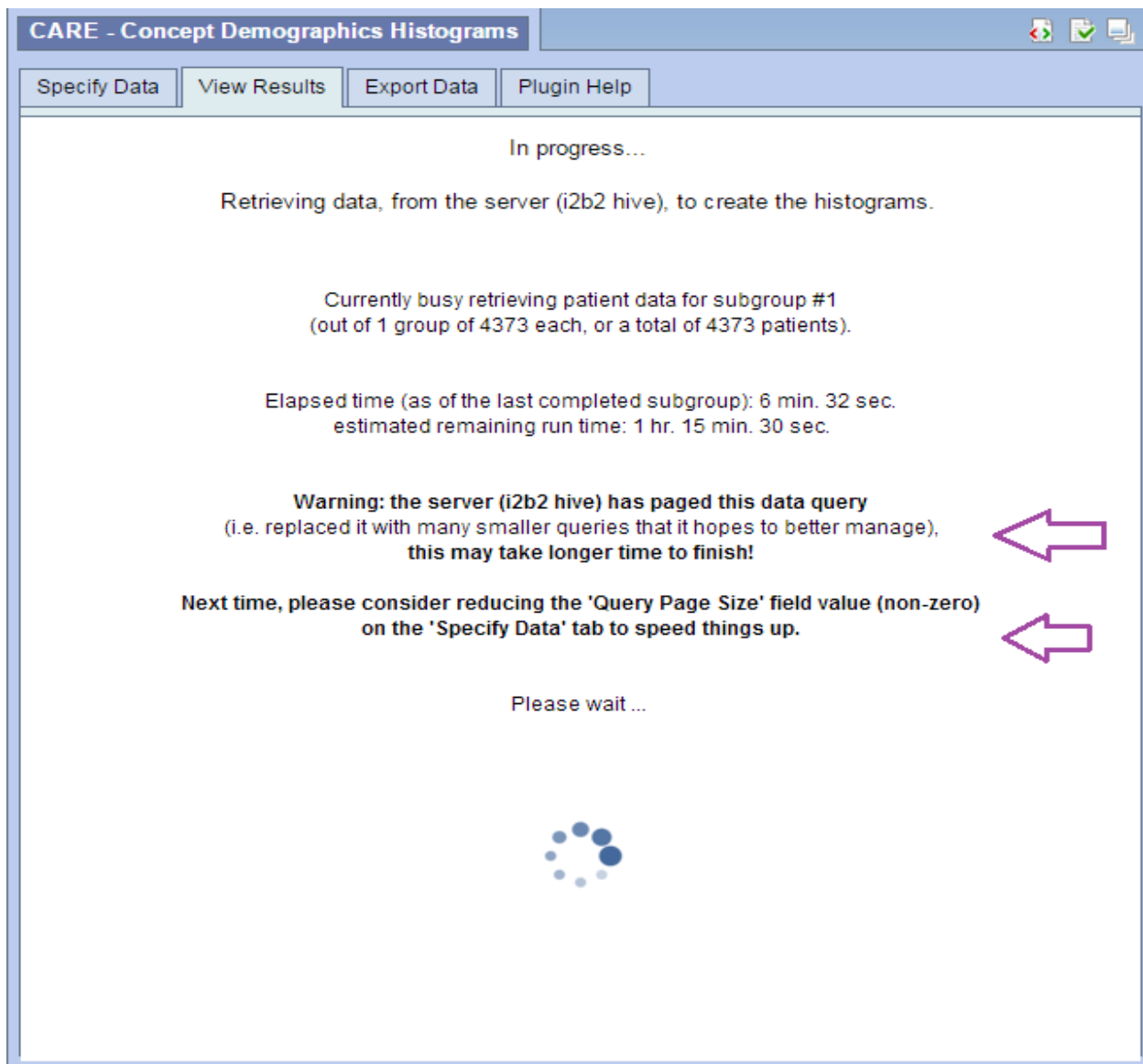
Notice that at the end of all the histograms are listed the total latencies of this plugin (one would need to scroll all the way down to the bottom of the display panel).



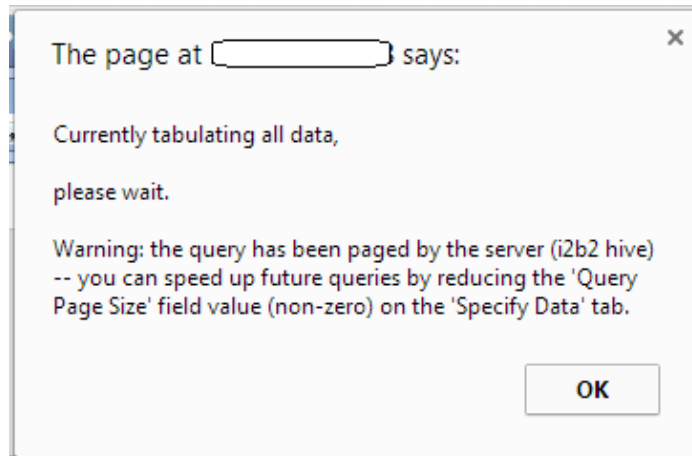
Warning About The i2b2 Hive Paging

When a user specifies very complex Concepts against very large Patient Set; the i2b2-hive may attempt to page that large query (or separate it into smaller sub-groups of queries, and then combine the results of these smaller queries into one set of data to be returned). However, sometimes it may lose track of its pages and cause an error resulting in an apparent hang in this plugin (i.e. this plugin appeared to be never finishing).

When it is detected that the i2b2-hive has paged the request, this plugin will display a warning (along with its estimates) message to urge the user to try options that would avoid future paging from occurring, like the following:



If the paging did not result in error, then a similar warning would also be displayed in the “all data retrieved” pop-up dialog, similar to the following (from a Chrome browser that was used) to again remind the user to try to avoid this same predicament in future runs (note: IP address blotched out):

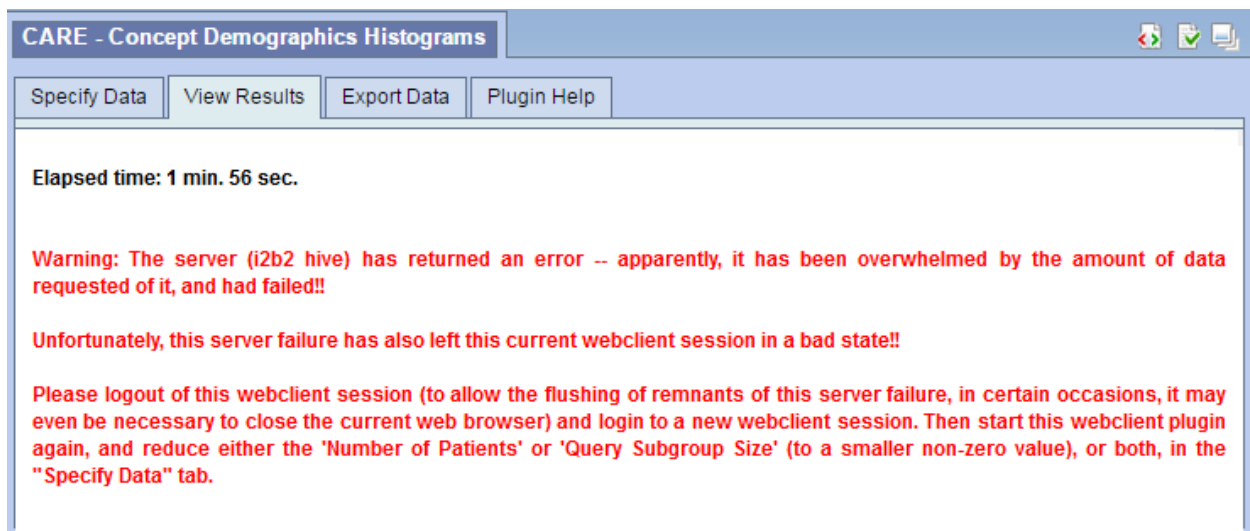


Unsuccessful Run

It is possible that when a user specifies many complex Concepts against huge Patient Set, a hang may result (i.e. this plugin may appear to never finish while its progress messages freeze).

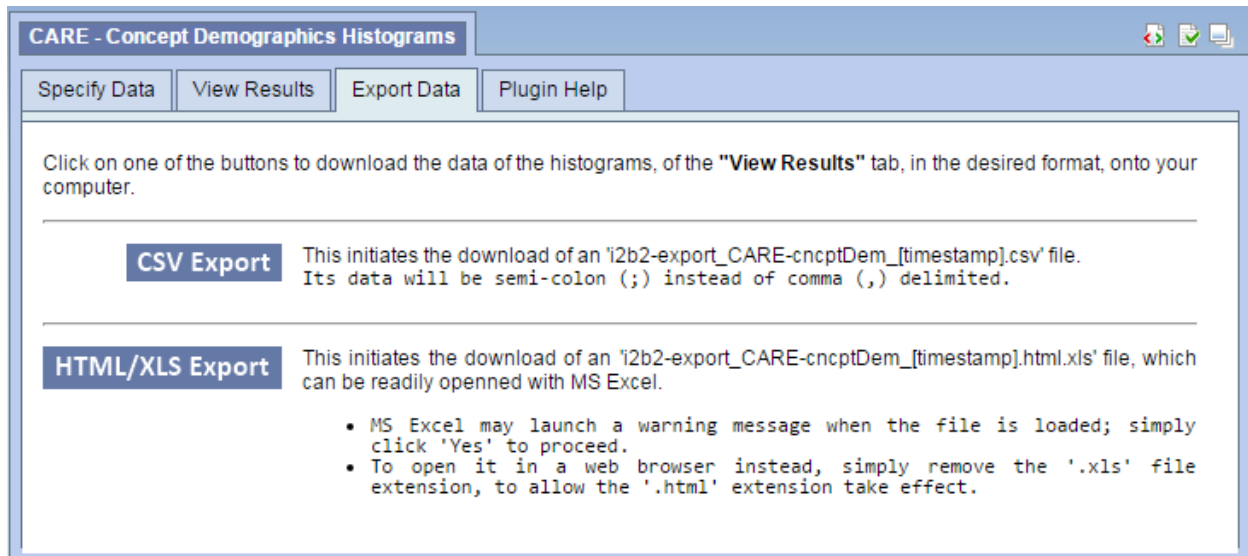
The cause of such situation has been traced to the i2b2-hive somehow losing track of its pages when it attempted to page the large query (or separate it into smaller sub-groups of queries, and then combine the results of these smaller queries into one set of data to be returned).

Despite of all the additional options introduced, it is still possible for one to have selected options that sensitize this problem of the i2b2-hive. When that occurs, then the following display will result:

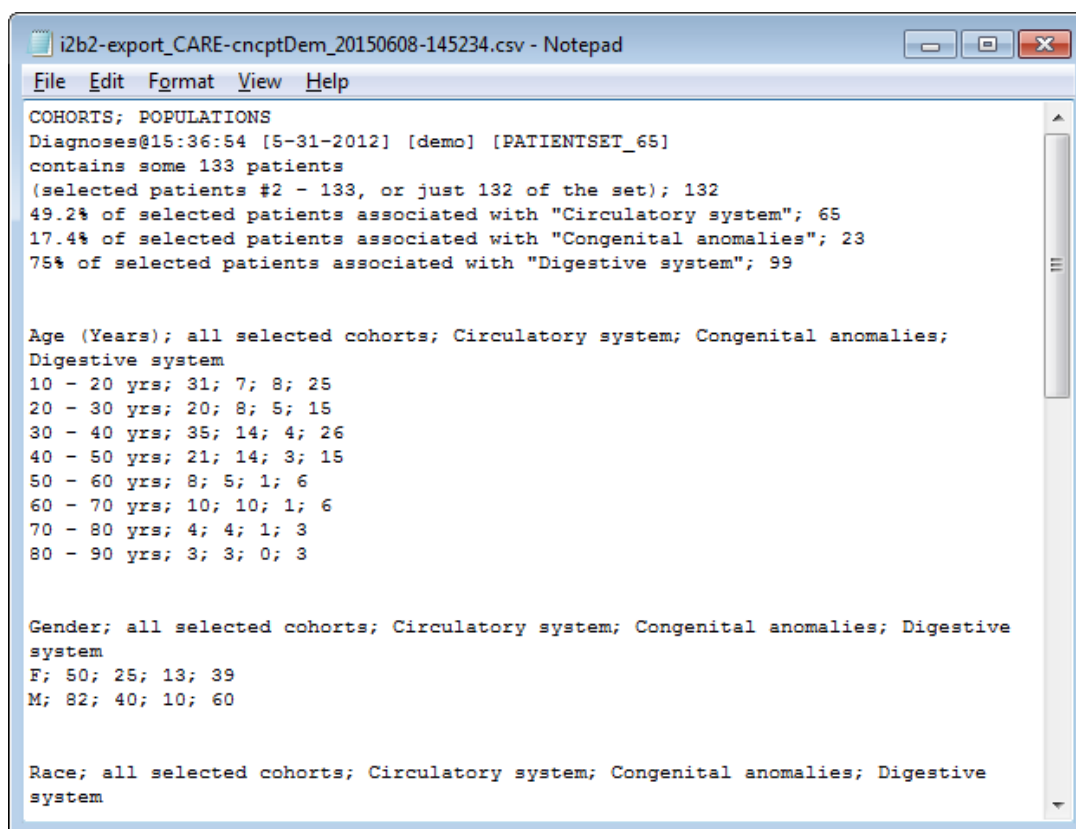


The “Export Data” Tab

Finally, select the "Export Data" tab to download the data of the histograms (in the “View Results” tab) to your computer.



The content of a sample exported CSV file is as following:



The content of a sample exported XLS file is as following:

COHORTS		POPULATIONS			
Diagnosed 15:36:54 15-31-2012 demo IPATIENTSET_851					
contains cases 133 patients					
(selected patients #2 - 133, or just 132 of the set)		132			
49.2% of selected patients associated with "Circulatory system"		65			
17.4% of selected patients associated with "Congenital anomalies"		23			
75% of selected patients associated with "Digestive system"		99			
Age (Years)		all selected cohorts	Circulatory system	Congenital anomalies	Digestive system
10 - 20 yrs		31	7	8	25
20 - 30 yrs		20	8	5	15
30 - 40 yrs		35	14	4	28
40 - 50 yrs		21	14	3	15
50 - 60 yrs		8	5	1	6
60 - 70 yrs		10	10	1	6
70 - 80 yrs		4	4	1	3
80 - 90 yrs		3	3	0	3
Gender		all selected cohorts	Circulatory system	Congenital anomalies	Digestive system
F		50	25	13	39
M		82	40	10	60

Caveats

Please note the following:

- Since it is possible for the server (i2b2 hive) to take a long time to provide all the data requested, this plugin has been enhanced with occasional updates of 'Elapsed time' and 'estimated remaining run time', etc. displays. These displayed values are at best rough estimates based on occasional data coming back from the server. In the case of zero 'Query Subgroup Size' selected (i.e. no query-subgrouping), then there would be no data coming back from the server until either of the following situations:
 - The whole, single, query is done.
 - The server itself started paging (i.e. query-subgrouping), and the result of the 1st of such paged subqueries just arrived.
 - The server got overwhelmed by all the excessive data and failed, returning an 'error'.

In any of these cases, the updates of 'Elapsed time' and 'estimated remaining run time' will be quite infrequent, as updates from the server (i2b2 hive) will most likely take a long time, and far and few in-between, if any.

- This plugin also tries to provide updates on the current subgroup the server (i2b2 hive) is fetching. However, in the cases when the server (i2b2 hive) itself started paging (i.e. query-subgrouping), then the current subgroup # being fetched may exceed the original stated total number of subgroups.

- The exact behaviors of exporting data in the desired files will be different among the various browsers (Firefox, Chrome, Internet Explorer, Safari, and Opera). Most of these browsers will present options between saving and just opening the exported file; while certain browsers, like Internet Explorer, may also open up a new blank tab page along with the export action.

Browser Compatibilities

- This plugin has been successfully tested with the following browsers: Firefox (v28, v31, v35), Chrome (v38, v40, v41), Internet Explorer 32/64 (v10, v11), Safari (v5.1.7) and Opera (v20, v27, v28); all on Windows 7.
- Its "Export Data" feature does not work on Vivaldi (v1.0.94.2).
- The i2b2 web client itself does not work with Opera (v12.11, v12.16).

i2b2 version compatibilities

This current version should be compatible with i2b2 v.1.6-1.7, but has been tailored to be compatible with v.1.7.0.7.

Version History

1.0	Initial release, with options of using a subset of a large patient set; user-friendly feedback (estimated run-time, etc.), helps & hints. By Wayne Chan, University of Massachusetts Medical School, Worcester, USA.	2015 Q1
1.0.1	Improved displays of the "COHORTS vs. POP. & LEGEND" table in the "View Results" tab on IE. By Wayne Chan, University of Massachusetts Medical School, Worcester, USA.	2015 Q2
1.1	Added "Export Data" options & other minor refinements. By Wayne Chan, University of Massachusetts Medical School, Worcester, USA.	2015 Q2
1.2	Fixed compatibility issue with webclient v.1.7.0.7, & other minor touch-ups. By Wayne Chan.	2015 Q4

Terms of Use

This plugin is published under the [GNU GPL v3 license](#).

This software is distributed as a plugin to the i2b2 v1.6-1.7 webclient framework, and may be used free of charge. It is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.