Data2Evidence Researcher Guide

Data2Evidence is an open source end-to-end platform for sharing healthcare data and analysing it.

User roles in Data2Evidence

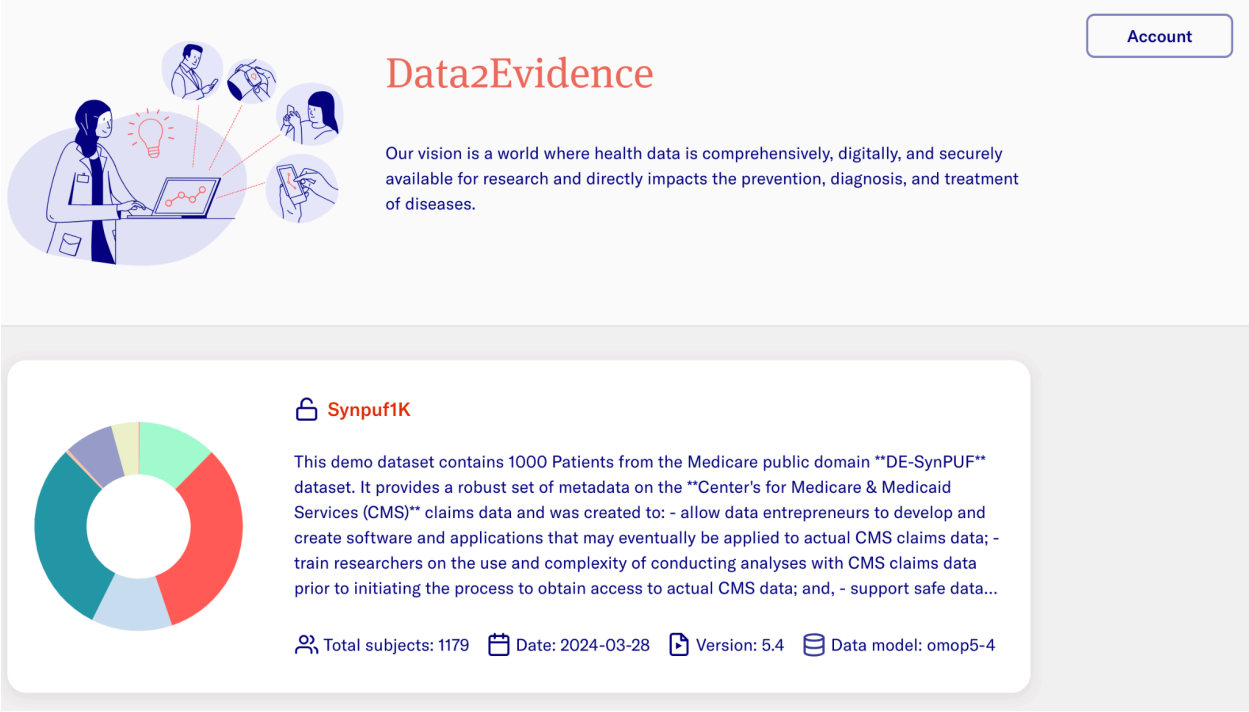
User roles determine the tasks you can perform when you sign in to the Data2Evidence platform. The key roles are:

- Viewer: Can view and request access to datasets.
- Researcher: Can view, build cohorts, and analyse datasets.
- Administrator: Can edit datasets, add users, and perform other administrative tasks.

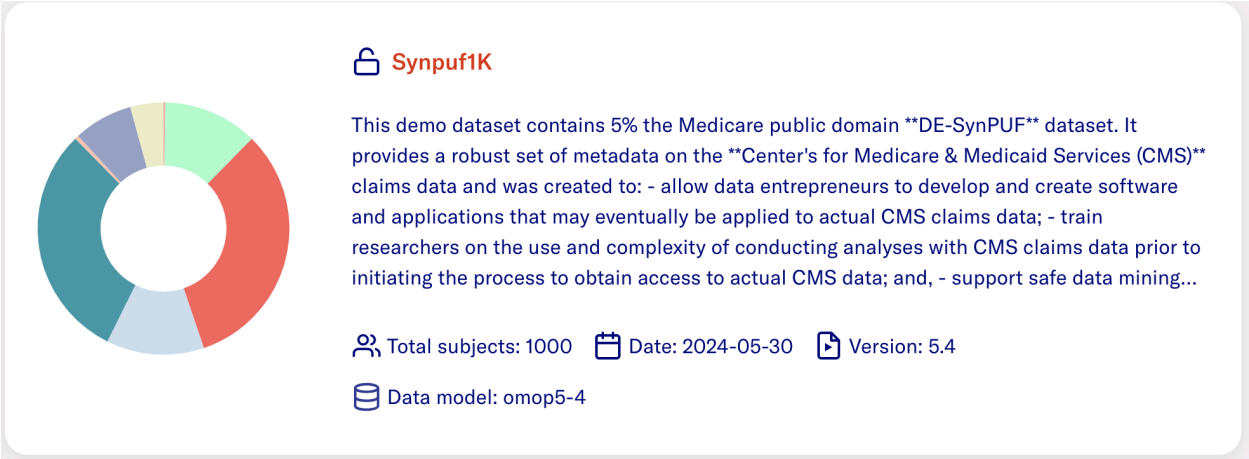
Overview of the user interface

After signing in, you see the dataset list page. Here, you can find:

- Publicly available datasets for everyone
- Private datasets to which you have been granted access



Dataset card



The dataset card here provides concise information about the dataset:

- Dataset name
- Short description
- More details like CDM version, patient count and create or update date

Click the card to proceed to the dataset page to perform further actions.

Navigation bar

The navigation bar on top has multiple options:

- **Dataset name dropdown:** See the current active dataset or use the dropdown to select another dataset
- **Dataset**: See current dataset information.
- **Concepts:** See the various concept sets available for the datasets or create a new concept set.
- **Cohort:** View the existing cohort list and access the cohort browser.
- **Notebooks:** Start the interactive notebook function.
- **Analysis:** Access pre-configured analysis pipelines
- **Account:** View your account details, change your password, read the legal documents, and sign out**.**

To go back to the dataset list page, click the Data2Evidence logo or the back arrow icon in the navigation bar.

Dataset

Dataset information page

The dataset information page contains further details about the dataset:

- Detailed information
- Details on how to access the data
- Metadata

- Download options for any resource files that the data provider added

If you only have the Viewer role, you can also request access from here using the Request access button. You can access the data quality dashboard and the data characterization dashboard from the tabs in this page.

Data quality dashboard

The data quality dashboard (DQD) executes a set of around 4,000 specific data quality checks against the database and assesses the results using a predetermined threshold. The quality checks are organized according to the Kahn framework. It employs a system of categories and contexts that stand in for methods for evaluating the quality of data. For full details on the DQD, you can read the official DQD documentation.

The DQD results table organizes the output according to the following main categories:

- **Plausibility:** Does the data conform with basic logical and medical expectations? *Example*: Does the measurement unit provided for a specific lab test make sense (for example, is it a unit like cm or m for body height)?
- **Conformance:** Does the data conform to the OMOP Common Data Model? *Example*: Is the patient ID given for a diagnosis entry indeed the primary ID of an entry in the PATIENT table?
- **Completeness:** Are all the expected data elements and vocabulary mappings present? *Example*: Does every medication entry have a standard OHDSI code identifying the medication given?

The DQD results table also provides a complete list of all checks run, including a check description, the fraction of failed rows, and the overall check pass/fail outcome.

Data characterization dashboard

The Data characterization dashboard (DCD) provides a top-level set of summary statistics to understand the data profile of a database in its totality. This quantitative assessment of a database typically includes questions such as:

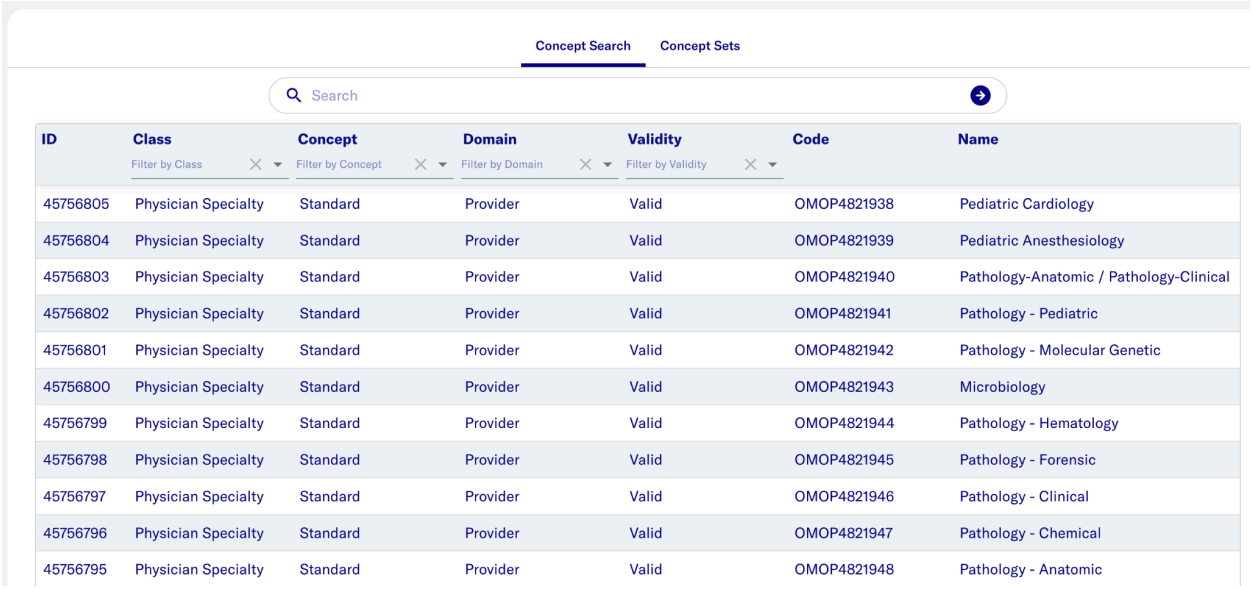
- What is the total count of persons in this database?
- What is the distribution of age for persons?
- How long are persons in this database observed for?
- What is the proportion of persons having a {treatment, condition, procedure, etc} recorded or prescribed over time?

These database-level descriptive statistics also help you to understand what data may be missing in a database.

To change the characterization detail, use the dropdown menu at the top. By default, it shows a summary characterization.

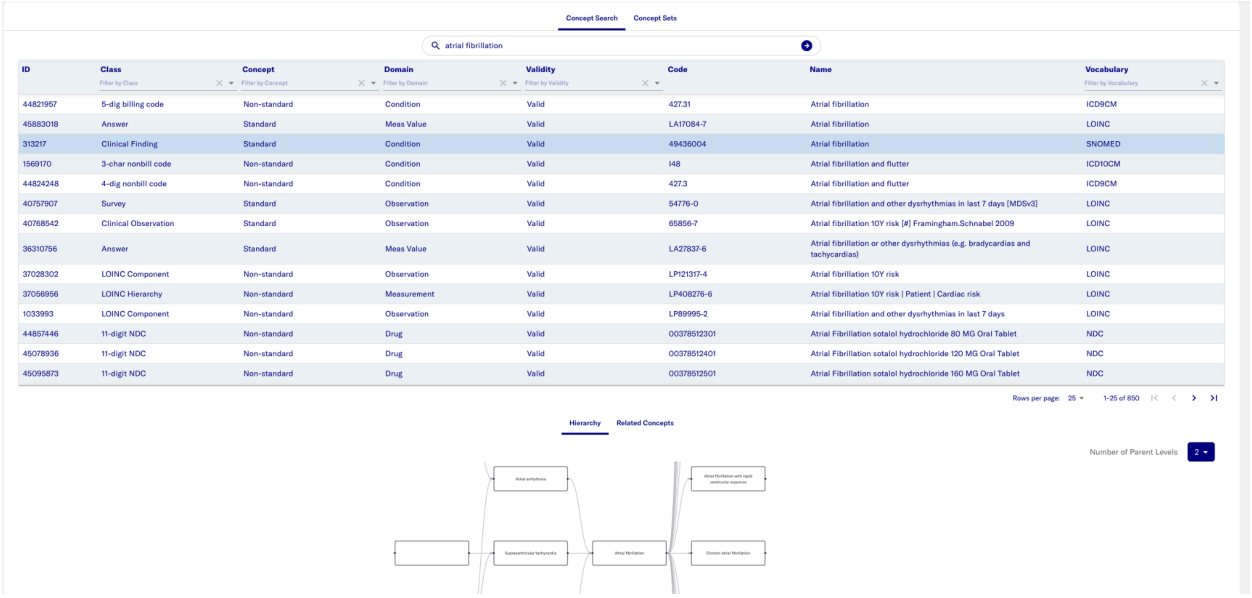
Concept sets

Concepts describe information in a patient's medical record, such as a condition, a prescription they are taking, a doctor's visit and other information. Domains are subject areas such as conditions, drugs, measurements, and so on. You can search for and save collections of concepts as a "concept set" and then use concept sets and cohorts to create a dataset, which you can use for analysis.



ID	Class	Concept	Domain	Validity	Code	Name
	Filter by Class	Filter by Concept	Filter by Domain	Filter by Validity		
45756805	Physician Specialty	Standard	Provider	Valid	OMOP4821938	Pediatric Cardiology
45756804	Physician Specialty	Standard	Provider	Valid	OMOP4821939	Pediatric Anesthesiology
45756803	Physician Specialty	Standard	Provider	Valid	OMOP4821940	Pathology-Anatomic / Pathology-Clinical
45756802	Physician Specialty	Standard	Provider	Valid	OMOP4821941	Pathology - Pediatric
45756801	Physician Specialty	Standard	Provider	Valid	OMOP4821942	Pathology - Molecular Genetic
45756800	Physician Specialty	Standard	Provider	Valid	OMOP4821943	Microbiology
45756799	Physician Specialty	Standard	Provider	Valid	OMOP4821944	Pathology - Hematology
45756798	Physician Specialty	Standard	Provider	Valid	OMOP4821945	Pathology - Forensic
45756797	Physician Specialty	Standard	Provider	Valid	OMOP4821946	Pathology - Clinical
45756796	Physician Specialty	Standard	Provider	Valid	OMOP4821947	Pathology - Chemical
45756795	Physician Specialty	Standard	Provider	Valid	OMOP4821948	Pathology - Anatomic

In the Concept Search tab, you can look for specific concepts and explore the hierarchy of concepts along with suggestions for other related concepts.

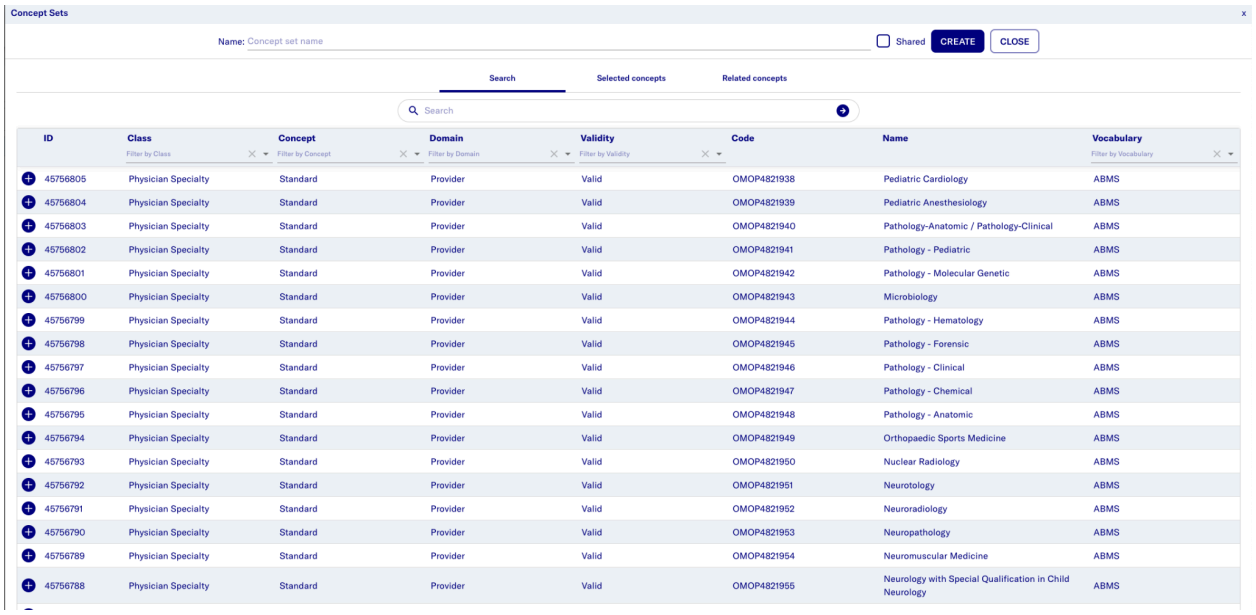


ID	Class	Concept	Domain	Validity	Code	Name	Vocabulary
44821957	5-dig billing code	Non-standard	Condition	Valid	427.31	Atrial fibrillation	ICD9CM
45883018	Answer	Standard	Meas Value	Valid	LA17084-7	Atrial fibrillation	LOINC
313217	Clinical Finding	Standard	Condition	Valid	49436004	Atrial fibrillation	SNOMED
1569170	3-char nonbill code	Non-standard	Condition	Valid	I48	Atrial fibrillation and flutter	ICD10CM
44824248	4-dig nonbill code	Non-standard	Condition	Valid	427.3	Atrial fibrillation and flutter	ICD9CM
40757907	Survey	Standard	Observation	Valid	54776-0	Atrial fibrillation and other dysrhythmias in last 7 days [MDSv3]	LOINC
40768542	Clinical Observation	Standard	Observation	Valid	65856-7	Atrial fibrillation 10Y risk [#] Framingham.Schnabel 2009	LOINC
36310756	Answer	Standard	Meas Value	Valid	LA27837-6	Atrial fibrillation or other dysrhythmias (e.g. bradycardias and tachycardias)	LOINC
37028302	LOINC Component	Non-standard	Observation	Valid	LP121317-4	Atrial fibrillation 10Y risk	LOINC
37056956	LOINC Hierarchy	Non-standard	Measurement	Valid	LP408276-6	Atrial fibrillation 10Y risk | Patient | Cardiac risk	LOINC
1033993	LOINC Component	Non-standard	Observation	Valid	LP89995-2	Atrial fibrillation and other dysrhythmias in last 7 days	LOINC
44857446	11-digit NDC	Non-standard	Drug	Valid	00378512301	Atrial Fibrillation sotalol hydrochloride 80 MG Oral Tablet	NDC
45078936	11-digit NDC	Non-standard	Drug	Valid	00378512401	Atrial Fibrillation sotalol hydrochloride 120 MG Oral Tablet	NDC
45095873	11-digit NDC	Non-standard	Drug	Valid	00378512501	Atrial Fibrillation sotalol hydrochloride 160 MG Oral Tablet	NDC

In the Concept Sets tab, you can search for any saved concept sets for a specific dataset. To edit any saved concept set, click the edit icon.

Creating a new concept set

1. Click “Add Concept Set” to start creating a new concept set.



ID	Class	Concept	Domain	Validity	Code	Name	Vocabulary
	Filter by Class	Filter by Concept	Filter by Domain	Filter by Validity			Filter by Vocabulary
45756805	Physician Specialty	Standard	Provider	Valid	OMOP4821938	Pediatric Cardiology	ABMS
45756804	Physician Specialty	Standard	Provider	Valid	OMOP4821939	Pediatric Anesthesiology	ABMS
45756803	Physician Specialty	Standard	Provider	Valid	OMOP4821940	Pathology-Anatomic / Pathology-Clinical	ABMS
45756802	Physician Specialty	Standard	Provider	Valid	OMOP4821941	Pathology - Pediatric	ABMS
45756801	Physician Specialty	Standard	Provider	Valid	OMOP4821942	Pathology - Molecular Genetic	ABMS
45756800	Physician Specialty	Standard	Provider	Valid	OMOP4821943	Microbiology	ABMS
45756799	Physician Specialty	Standard	Provider	Valid	OMOP4821944	Pathology - Hematology	ABMS
45756798	Physician Specialty	Standard	Provider	Valid	OMOP4821945	Pathology - Forensic	ABMS
45756797	Physician Specialty	Standard	Provider	Valid	OMOP4821946	Pathology - Clinical	ABMS
45756796	Physician Specialty	Standard	Provider	Valid	OMOP4821947	Pathology - Chemical	ABMS
45756795	Physician Specialty	Standard	Provider	Valid	OMOP4821948	Pathology - Anatomic	ABMS
45756794	Physician Specialty	Standard	Provider	Valid	OMOP4821949	Orthopaedic Sports Medicine	ABMS
45756793	Physician Specialty	Standard	Provider	Valid	OMOP4821950	Nuclear Radiology	ABMS
45756792	Physician Specialty	Standard	Provider	Valid	OMOP4821951	Neurotology	ABMS
45756791	Physician Specialty	Standard	Provider	Valid	OMOP4821952	Neuroradiology	ABMS
45756790	Physician Specialty	Standard	Provider	Valid	OMOP4821953	Neuropathology	ABMS
45756789	Physician Specialty	Standard	Provider	Valid	OMOP4821954	Neuromuscular Medicine	ABMS
45756788	Physician Specialty	Standard	Provider	Valid	OMOP4821955	Neurology with Special Qualification in Child Neurology	ABMS

2. To search for any concept across all domains, use the search field. Click the plus icon next to the ID to add the concept to your set. You can filter by
- class,
- concept (standard/non-standard),
- Domain, or
- validity (valid/invalid).

   You can see the concepts added to the set by looking at the Selected concepts tab and can click on the minus sign to remove any you might not want. You can also look into the related concepts tab to get more suggestions on concepts that might be relevant to your concept set based on the concepts you have already selected.
3. Enter a concept set name and click on create to save the concept. Any changes made after this can be saved by clicking on Update.
4. Click Close to close this window.

All the saved concept sets can be used in the cohort creator to create cohorts.

Cohorts

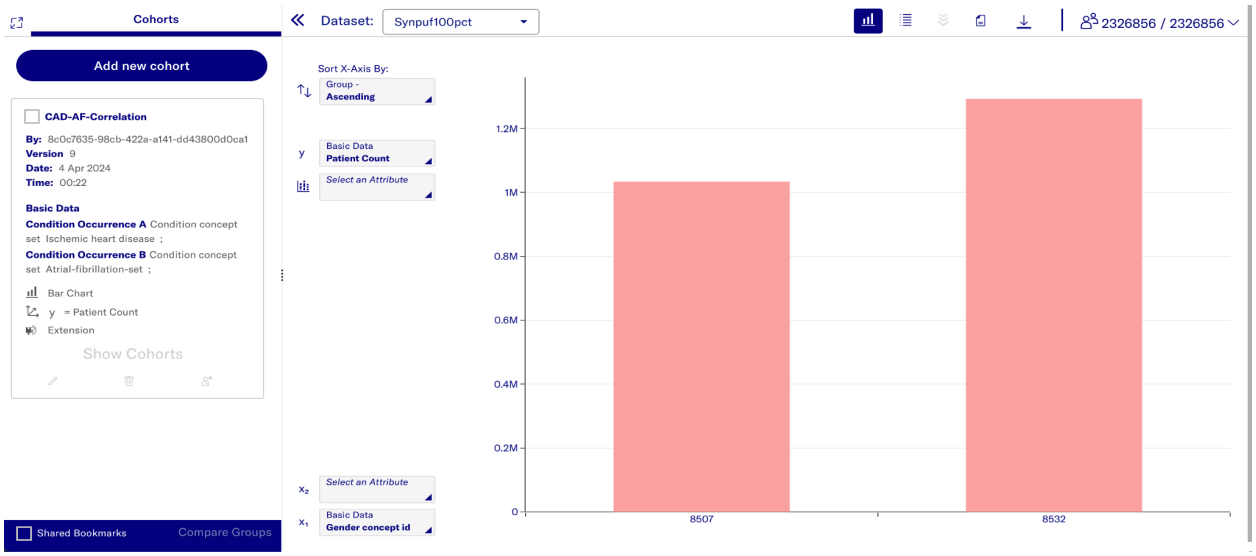
Cohorts (also called phenotypes) are a set of persons that fulfil some inclusion and exclusion criteria for a duration of time. A cohort can also be defined independently of the other cohorts used in a study, thereby enabling their reuse for other studies. Data2Evidence enables building of cohorts using a cohort creator that provides instantaneous feedback on cohort sizes and other parameters.

Cohort overview

You can see a list of cohorts you have already created and saved in the cohort overview. To see cohorts that were shared with you, select the Shared Bookmarks checkbox on the bottom left of the page.

Cohort creator

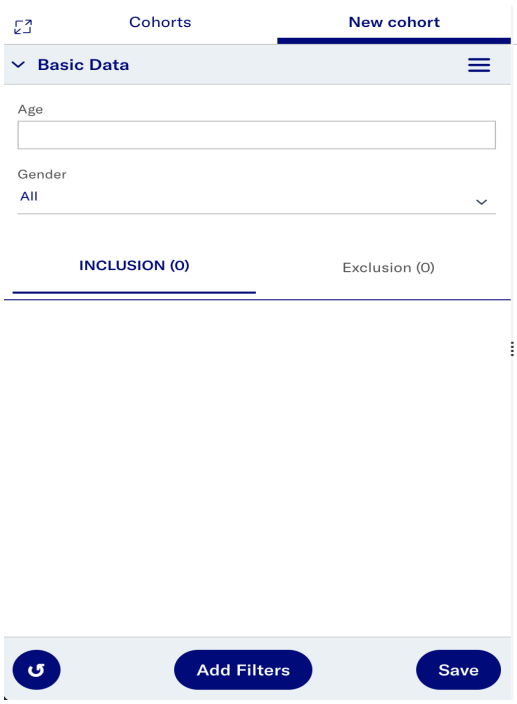
The cohort creator (previously known as patient analytics) is a no-code, browser-based, interactive solution to build cohorts. You can access it by clicking on the cohort tab in the navigation bar.



To load a cohort, click any of the cohort cards. You can click on Add new cohort to start the cohort creation process. You need to verify and select the relevant dataset from the dropdown list on top before the creation of a new cohort.

Filter area

The filter area is on the left of the screen. Each filter is shown in this area as a filter card. This is also where you specify and refine the filter criteria for creating your cohorts by adding filter cards. The default filters loaded depend on the installation.



The filter cards in the Inclusion tab are the inclusion criteria and persons matching the criteria here are included in the cohort and the visualisations. The filter cards on the Exclusion tab are the exclusion criteria and persons matching the criteria here are excluded from the cohort and the visualisations.

You can add new filter cards by clicking on the Add filters button. This will show a menu of the kinds of filter

cards that can be added. You can choose the domain of the filter card and subsequently, choose the criteria for the card. The default fields are concept set and domain type concept set. To add other criteria to this card, click the menu icon (3 lines) at the top right of the card. You can remove, reset or rename the filter card from this menu too.

You can enter your search terms in the fields of the filter cards and the system will display a list of suggestions as you type. The number of values for any attribute displayed in the dropdown list is configured by the system administrator. If the dropdown displays the message: “Refine your search”, the number of suggestions exceeds the configured system limit. Continue entering text to further refine your search and see the suggestions displayed. If no suggestions are available, the message is “No suggestions available”.

If you add a new filter card after the first one has been configured, you will get the “AND” boolean relationship between the subsequent cards. This means that persons fulfilling both filter card criteria will be included/excluded from the cohort. You can click this “AND” button to switch to the “OR” relationship which means that persons fitting either filter card criteria will be included/excluded from the cohort.

If the filtering criteria does not have any relevant persons in the dataset or the number of persons in the cohort drops below a certain threshold, a warning message is displayed on the top of the screen.

To add a temporal relationship between any filter cards, add the advanced time option from the Menu of one of the filter cards. Define the time interaction in terms of days, and in relation to the start and end dates of the interactions. There can also be an overlapping temporal relationship.

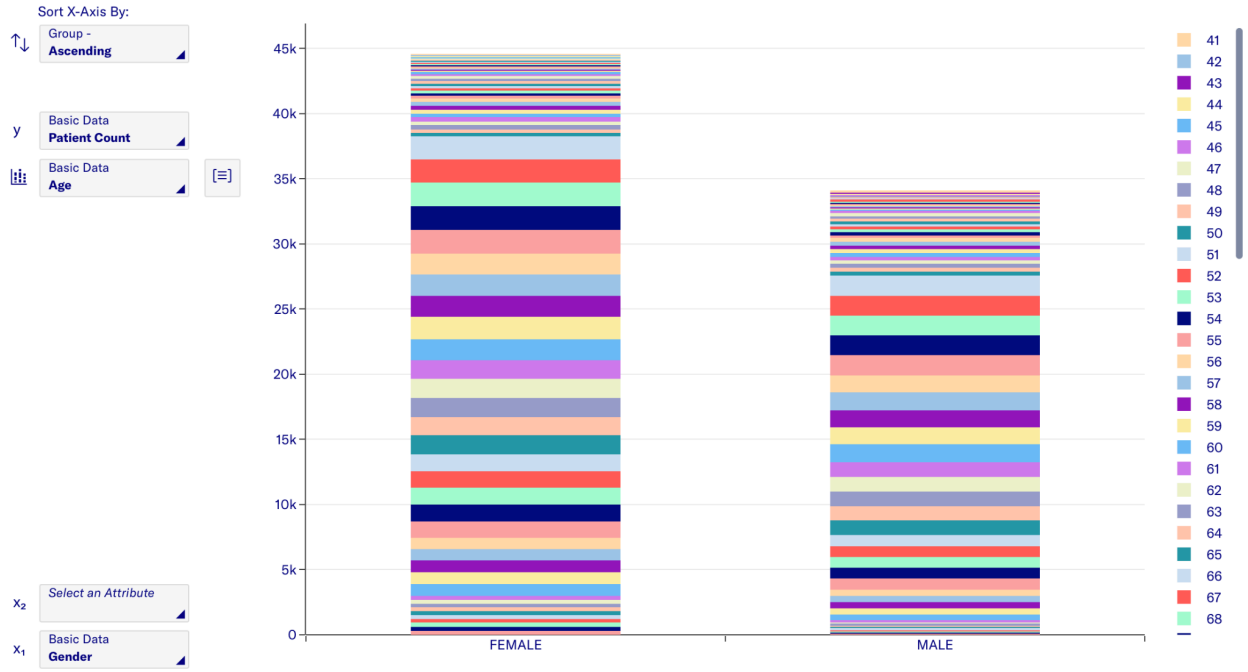
To save your filter cards, click Save and enter the details. This filter setting can also be shared with other users. Clicking the Reset icon resets everything in the cohort creator, including filter cards and previously added charts.

If you are working with a lot of filter cards and want to maximise the space available to the filtering area, click the “Enter Fullscreen” icon on the top left. This will hide everything else other than the filtering area. To return to the original size, click the “Exit Fullscreen” icon on the top left.

Chart area

The chart area is on the right of the screen. It displays a bar chart of the information, by default. It changes continuously based on the filters applied.

Move your cursor over the column to get a quick view of the attributes that form the group and its numeric value. You can also choose a second attribute for the x-axis to split the columns or for the y-axis to have a stacked chart. You can reset any of the selections using the dropdown menu.



Charts are only displayed for groups with a minimum number of persons. Results with fewer persons than the specified minimum are not displayed on the charts. This ist to preserve privacy and stops you from being able to recognize individual persons based on the charts.

You can also click on a column in the graph and click the “Filter by Selection” button to filter the data and add the attributes of the column to the filter cards.

You can change the sorting of the groups on your x-axis by choosing the order in the dropdown of the x-axis. The default sorting is either alphabetical or numerical in ascending order. You can also reduce the number of groups by combining together groups within a range of numerical attributes. This is called binning. Use the binning icon next to the axis selector field and specify the bin size.

You can export the graph as a CSV or PNG file by using the download option. These exports do not contain your filter information or underlying data.

It is possible to view the patient list by clicking on the Patient list button, if it has been enabled. You might not be able to see all the patients that match your filter criteria in the patient list. The patient list doesn't show aggregated data – so you can only see the patients that you're allowed to access. You can edit this patient list table to include additional criteria and interactions to show. You can also download this table, if downloading has been enabled by your administrator.

Additional options

You can view a summary of the current filter criteria by clicking on the filter summary option. The number on the top right of the screen is the current person count shown as the number of

persons fulfilling the filtering criteria or total number of persons in the database. This may be a restricted number of persons, depending on your access rights.

Notebooks

You can click the Notebooks option to start Starboard Notebooks. These are in-browser, literate notebooks. You can use Python code here to analyse your cohorts or the dataset. You can:

- Create a new notebook
- Import an existing notebook
- Access other shared notebooks from here

Note: Starboard notebooks are not meant for intense processing and analysis of large datasets. You should use your own environment to do that.

Analysis

The drop down here provides a list of different pre-configured analyses that can be performed on your dataset. At this point of time, we have Kaplan Meier Survival Analysis.

Account details

You can click the Account tab to access

- Details about your account
- Legal documents (Terms of Use, Privacy Policy, and Imprint)
- Change your password
- Delete your account
- Sign out from the system.

Change password

Click on change password to change your password. Enter your old password and your desired new password in the fields provided. You can also get the system to generate a random password for you. Click on update to save the new password or click on cancel to cancel the change process.