Instructions for using Mica Search Tool of euCanSHare Data Catalogue

www.eucanshare.eu

4th November 2022, Version 1

Tarja Palosaari and Jaakko Reinikainen, Finnish Institute for Health and Welfare (THL)

Contents

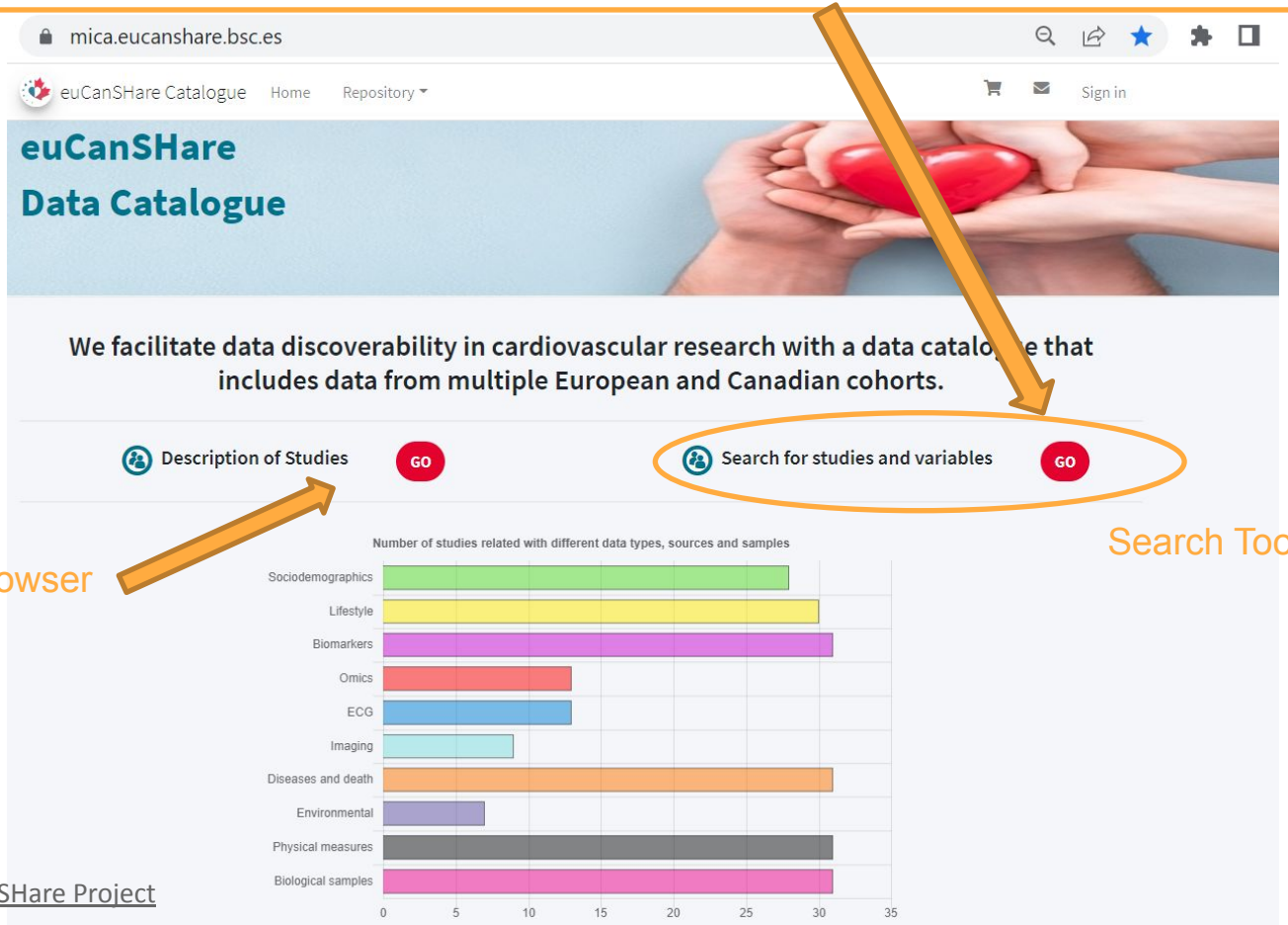
euCanSHare Data Catalogue

- Data Catalogue is a discovery tool for cardiovascular research data.
- It provides detailed information on
  - characteristics of the studies, such as study design and data access policies;
  - data collected in the studies, such as variable descriptions;
  - omics and bioimaging data available for the studies.
- Data Catalogue is built using OBiBa's Mica and Opal software applications, including:
  - Cohort Browser that provides a structured description of the participating studies, and
  - Search Tool that allows users to browse information within the Data Catalogue using a powerful search engine.
    - Instructions for using the Search Tool are provided in this tutorial.

More information on catalogue tools

- OBiBa tools are open source software for data management, harmonization, co-analysis and dissemination of the epidemiological research.
- OBiBa tools are described in detail in the websites of the Maelstrom Research and OBiBa
- Watch also the video tutorial on using the search of the Maelstrom Catalogue that is built on same tools as the euCanSHare Data Catalogue
    - Video: https://www.maelstrom-research.org/page/tutorials?topic=search

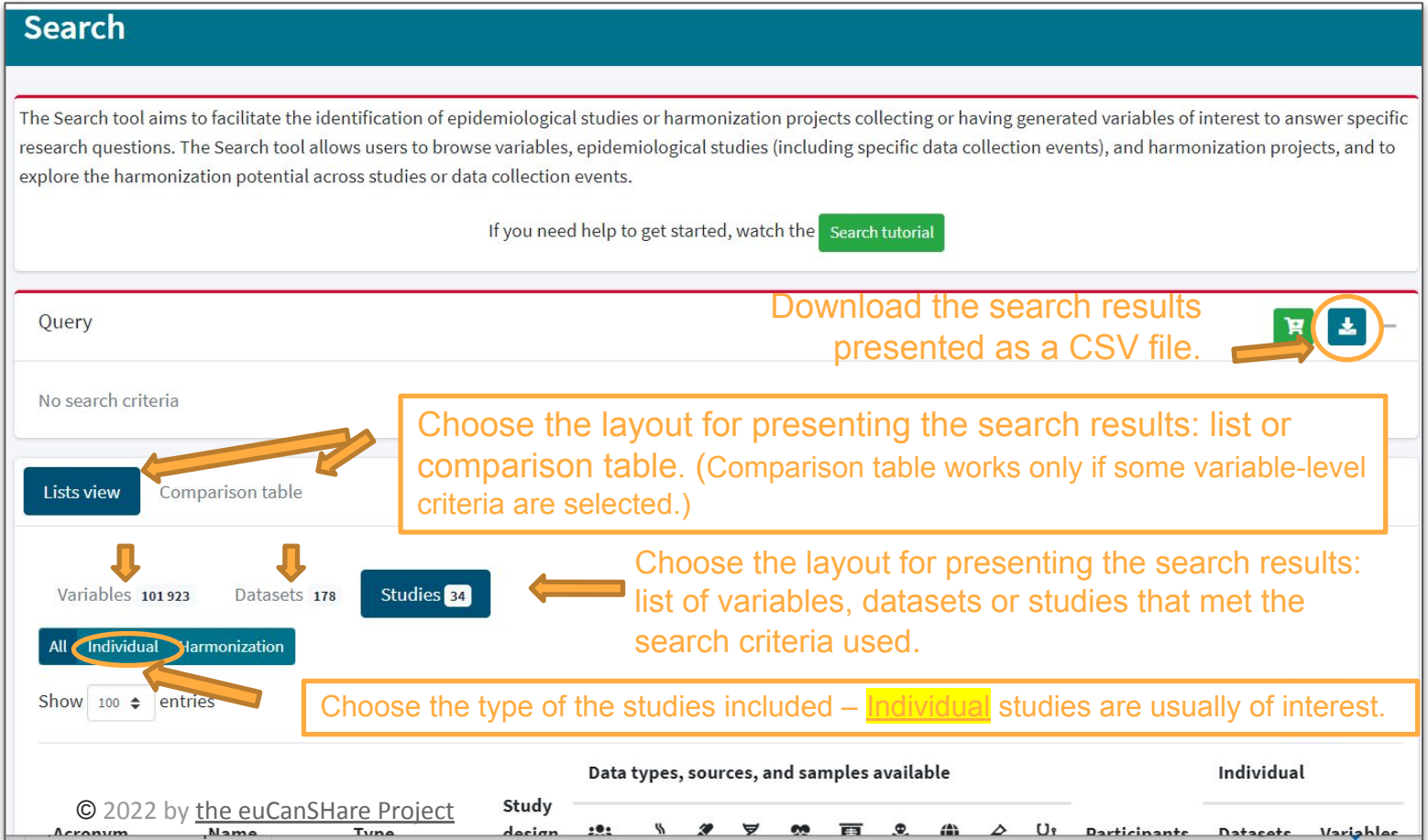
euCanSHare Data Catalogue frontpage: https://mica.eucanshare.bsc.es/
Mica Search Tool can be accessed by clicking GO in "Search for studies and variables"



Cohort Browser

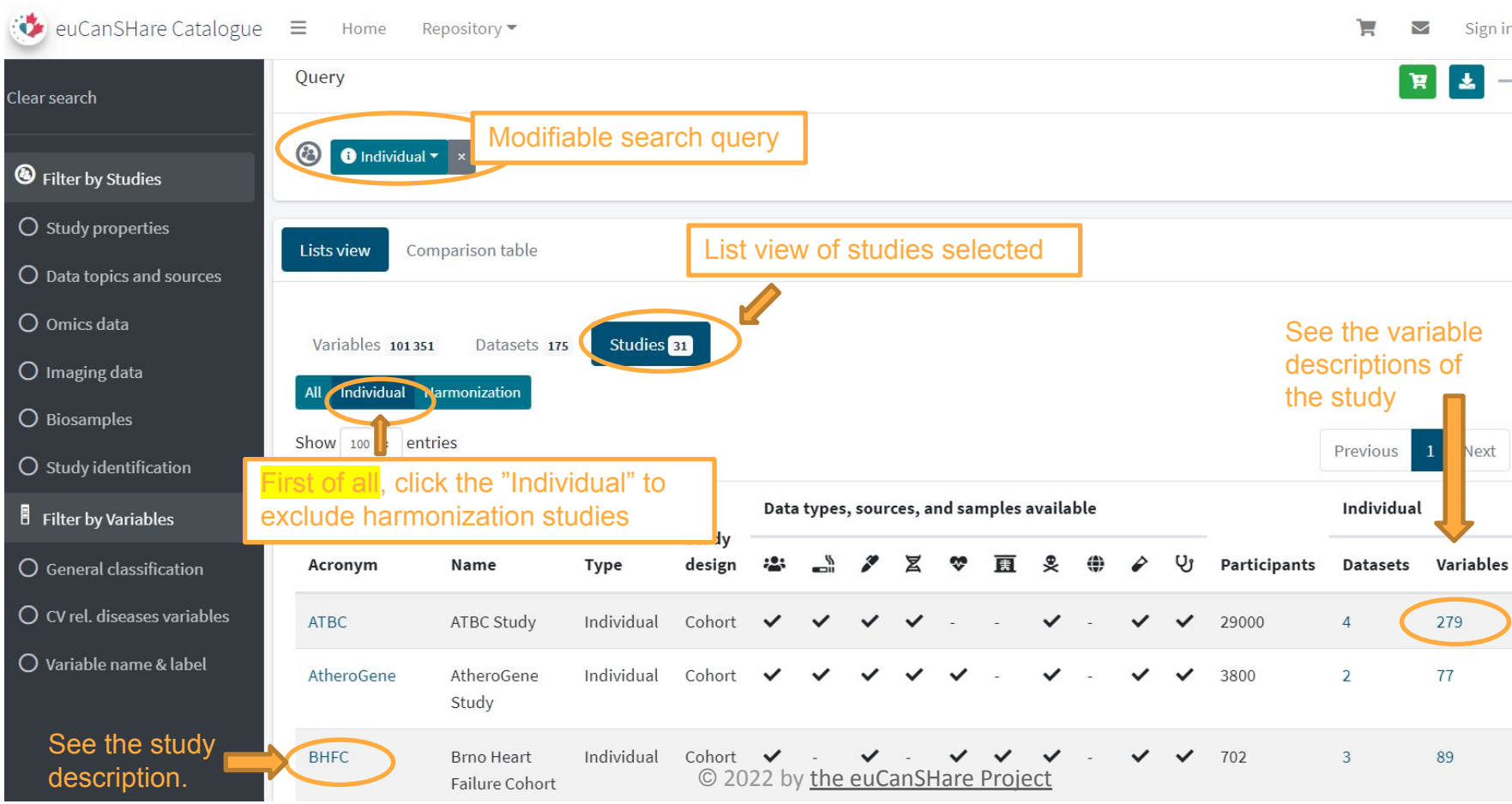
Search Tool

Functionalities of the Search Tool (1/22)
Basic functionalities

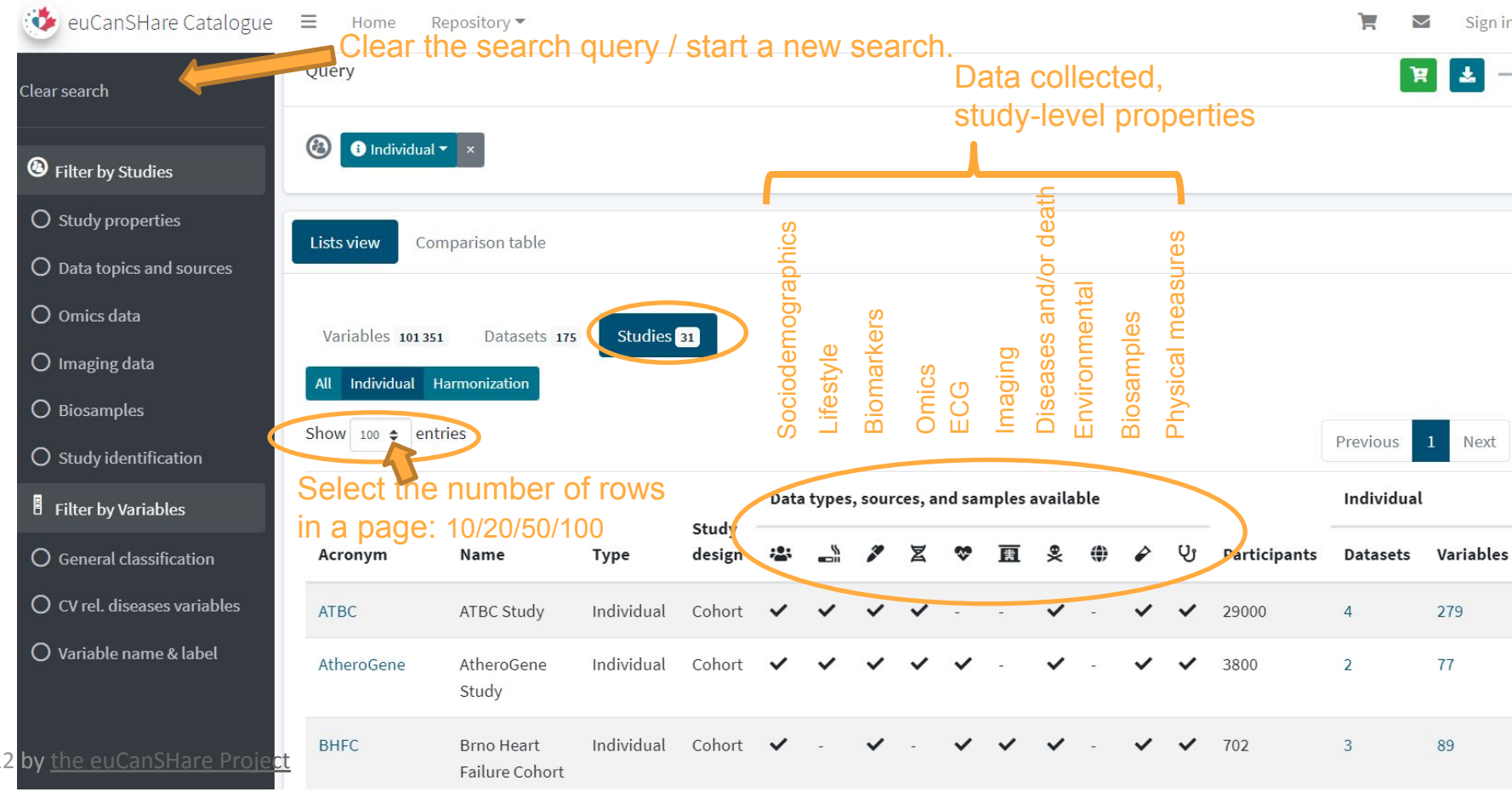


A **harmonization study** is a research project harmonizing data across individual studies to answer specific research questions.

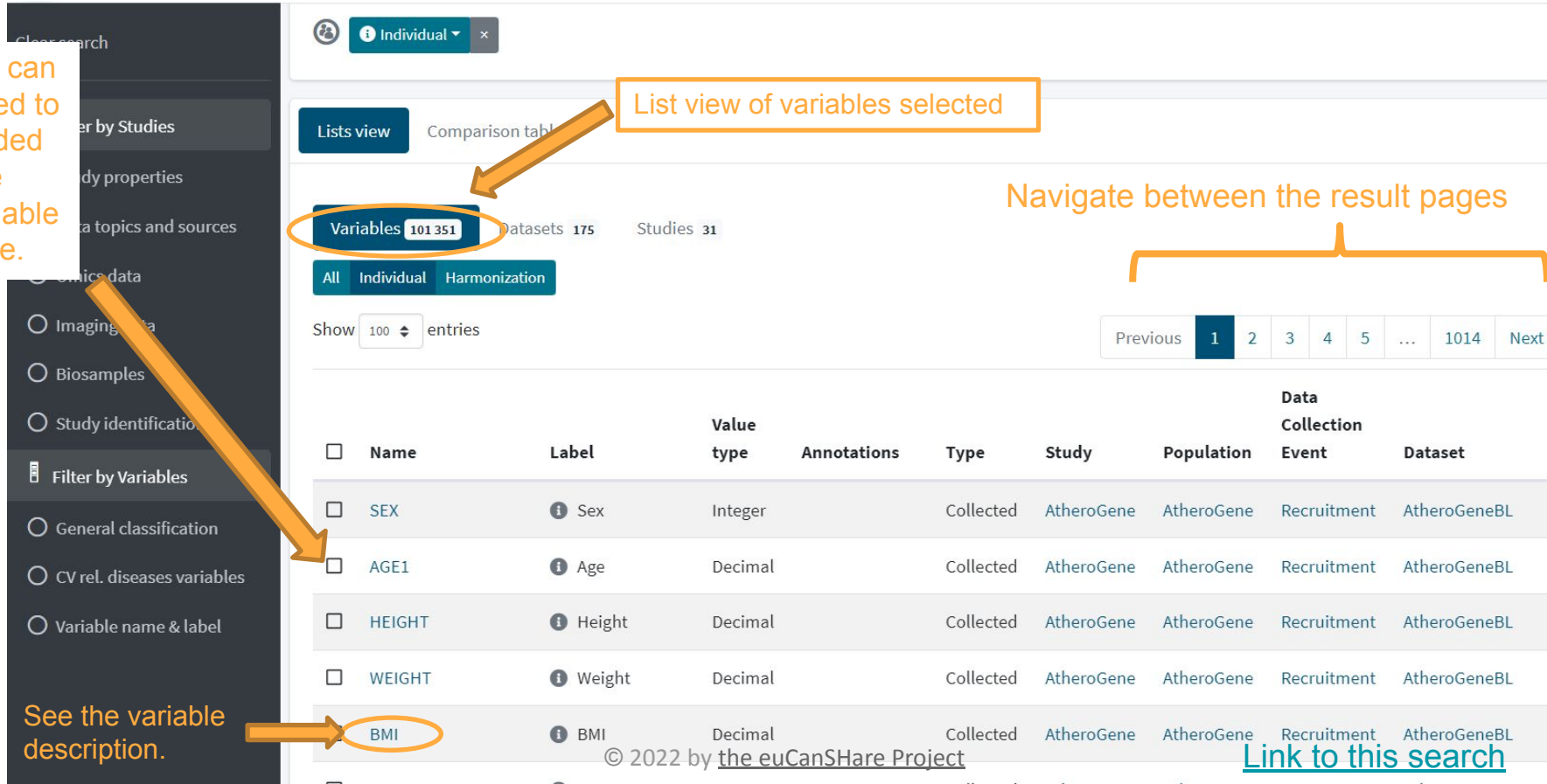
Functionalities of the Search Tool (2/22)
Basic functionalities - Results as a list of studies



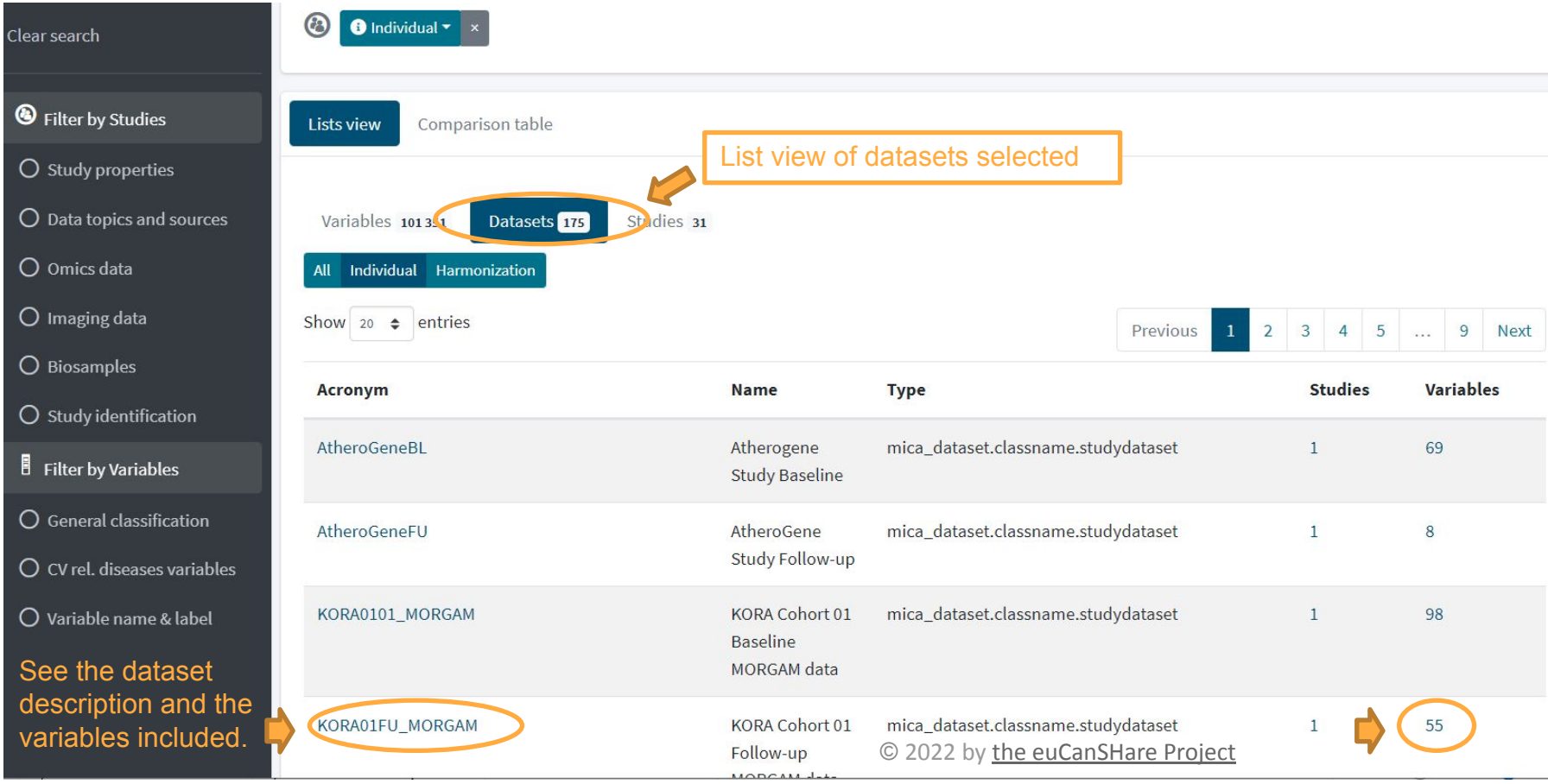
Functionalities of the Search Tool (3/22)
Basic functionalities - Icons for data collected in the studies



Functionalities of the Search Tool (4/22)
Basic functionalities - Results as a list of variables



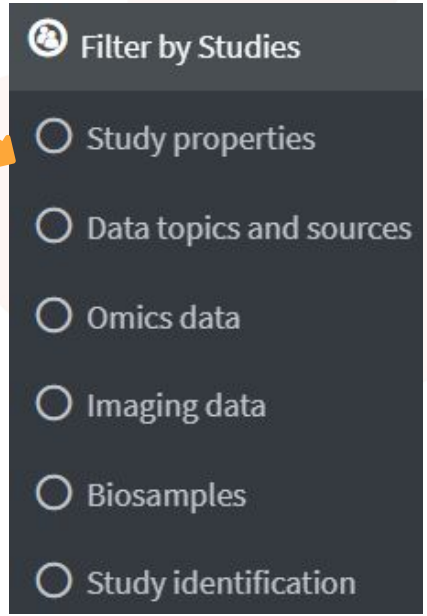
Functionalities of the Search Tool (5/22)
Basic functionalities - Results as a list of datasets



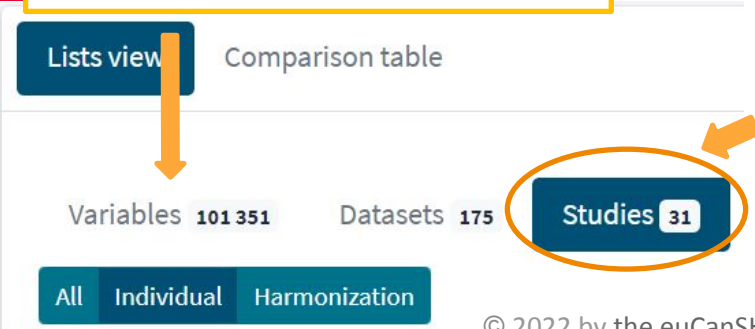
Adding search criteria and presenting results (1/2)

Search criteria can be applied either using properties defined on 1) study-level or 2) variable-level, or both, and search results in all cases can be chosen as list of studies, datasets or variables.

1) To search by using study-level properties i.e. filtering search results based on data collected in the studies or the features of the participants. “Filter by Studies”



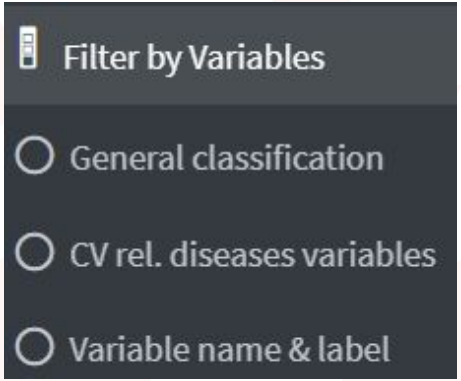
Includes ALL variables in the studies.



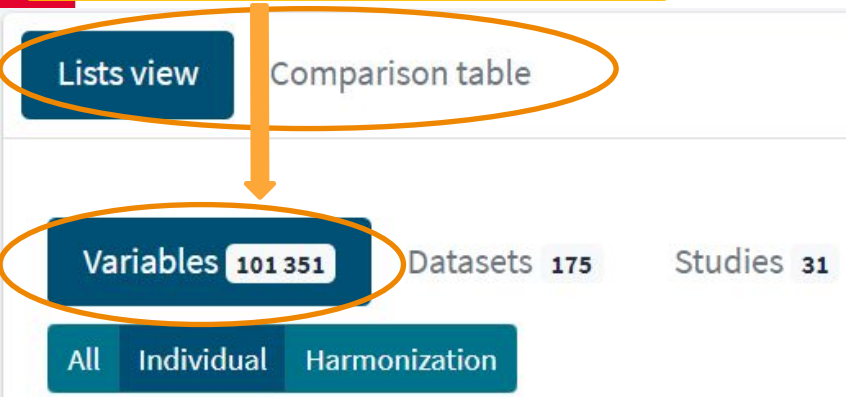
Note that filtering on study-level affects only to the studies included in the results, and NOT the variables by their content.

Adding search criteria and presenting results (2/2)

2) To search by using variable-level properties i.e. filtering search results by the contents of the variables included in the studies. “Filter by Variables”



Variables will be filtered by their content when variable-level criteria are used.

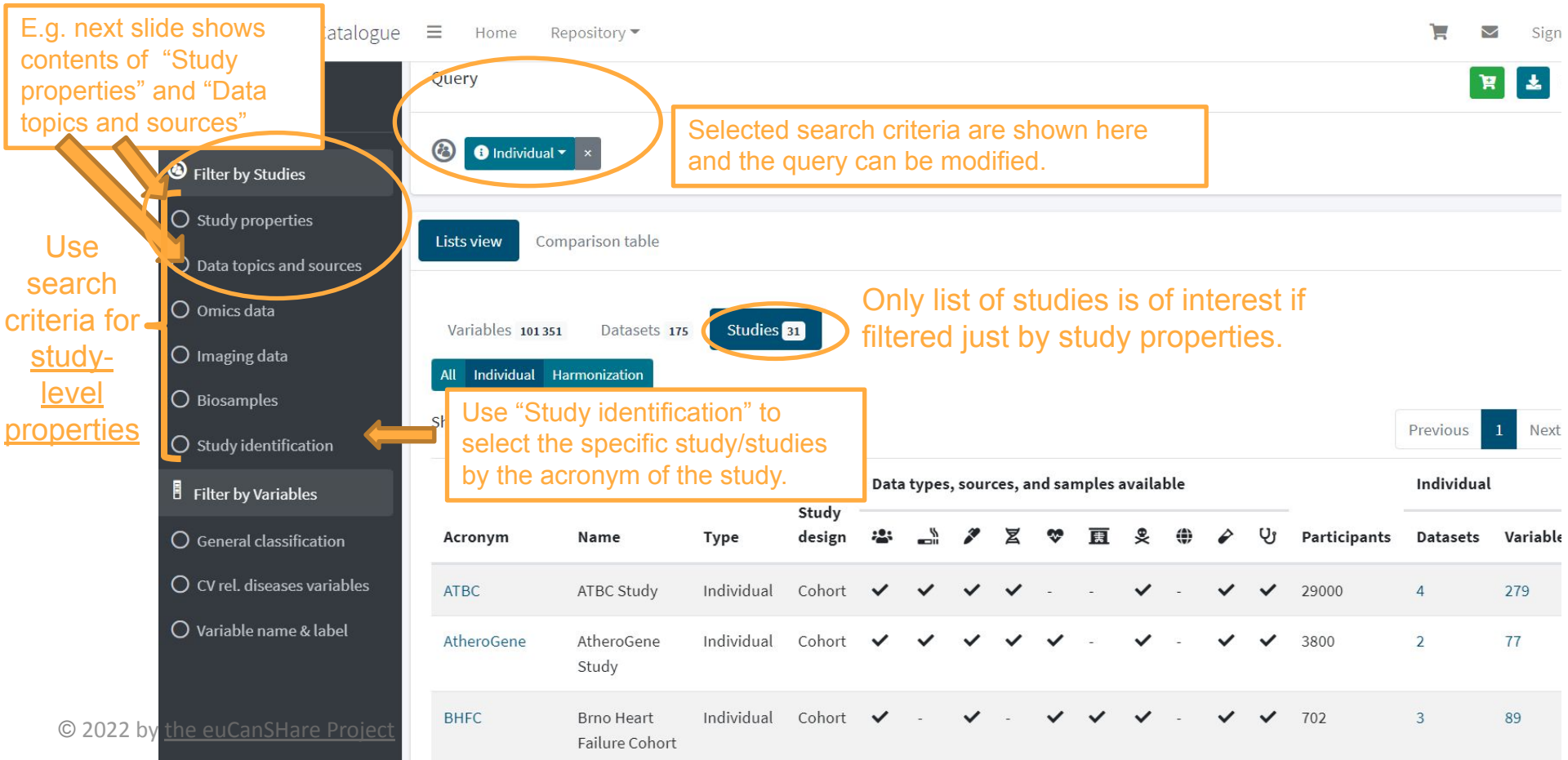


Filtering on variable-level affects in addition to studies included, also to the list of the variables by their content. Comparison table view can also be selected after filtering by variables’ content.

Functionalities of the Search Tool (6/22)
1) Adding search criteria in study-level - Filter by studies

E.g. next slide shows contents of "Study properties" and "Data topics and sources"

Use search criteria for study-level properties



Functionalities of the Search Tool (7/22)
1) Adding search criteria in study-level - Study properties

Study properties

Study properties as defined in the catalogue.

Here selected are "cohort" from "study design" and "general population" from "sources of recruitment"

Select all subcategories at once

Study design — Select All | Clear Selection

The design of an observational or experimental study (e.g. cohort, case control).

- [x] Cohort
- [ ] Case-control
- [ ] Case only
- [ ] Cross-sectional
- [ ] Clinical trial
- [ ] Other

▾ More

Observational study that involves the analysis of data collected from a population at one specific point in time.

Pointing by mouse gives the description of the category.

Selection criteria - Country of residence — Select All

Participant's country of residence.

- [ ] Andorra
- [ ] United Arab Emirates
- [ ] Afghanistan
- [ ] Antigua and
- [ ] Anguilla
- [ ] Albania
- [ ] Armenia
- [ ] Angola
- [ ] Antarctica
- [ ] Argentina
- [ ] American Samoa
- [ ] Austria
- [ ] Australia
- [ ] Aruba
- [ ] Åland Islands
- [ ] Azerbaijan
- [ ] Bosnia and Herzegovina
- [ ] Barbados
- [ ] Bangladesh
- [ ] Belgium
- [ ] Burkina Faso
- [ ] Bulgaria
- [ ] Bahrain
- [ ] Burundi
- [ ] Benin

More: See all categories if the list is long ▾ More

Then click: Display results

Selection criteria - Sex — Select All

Participant's sex

- [ ] Men only
- [ ] Women only

▾ More

Sources of recruitment — Select All | Clear Selection

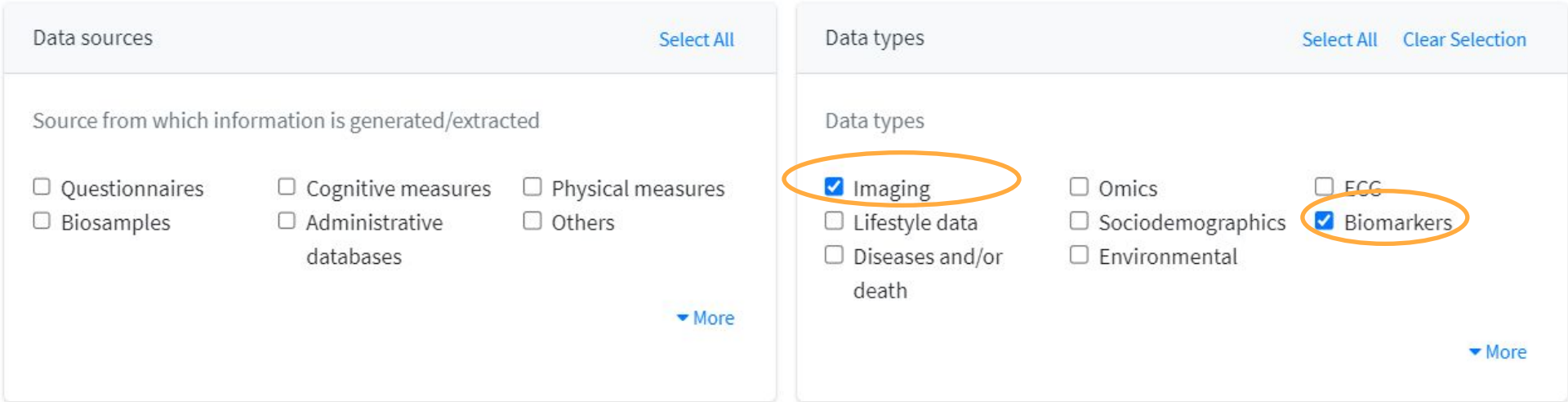
The population(s) from which individuals are recruited to participate in the study.

- [x] General population
- [ ] Specific population
- [ ] Participants from existing studies

Functionalities of the Search Tool (8/22)
1) Adding search criteria in study-level - Data topics and sources

Here selected are “imaging” and “biomarkers” from “data types”

Click: Display results



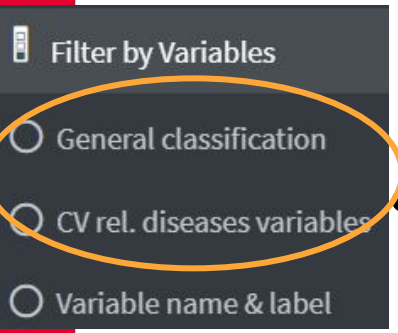
Functionalities of the Search Tool (9/22)
Results by studies when some study-level properties are selected.
Modifying the search query

By clicking small arrows in the query you can modify the criteria. Note that OR "|" is the default operator when selecting many categories within one section; AND-operator can not be applied here.

Operation	Description
any	Classified to any non-empty category.
none	Classification must be missing.
in (default)	Classified to at least one of the selected categories.
not in	The negation (opposite) of "in".

Link to this search

Filter by variables -options (1/2)



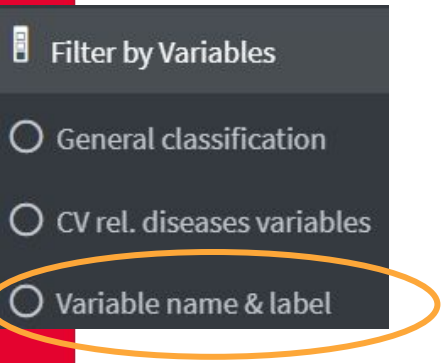
- General classification
  - "Areas of Information" classification for the topics of the variables, developed by Maelstrom Research
  - A variable is classified into some subcategory of e.g. "Lifestyle and behaviours", "Diseases" or "Laboratory measures"
- Cardiovascular (CV) related disease variables
  - This classification includes some specific CV related diseases, that are subcategories of broader definition of the diseases, developed by euCanSHare

These classifications make it possible and easy to search for variables in the specified areas of interest across the studies.

- Studies must have added the variable descriptions and also classified each variable into these categories, to be able to find that study and its variables when using variable-level filtering.

These two variable classifications are the recommended way to perform a search using the variable properties.

Filter by variables -options (2/2)

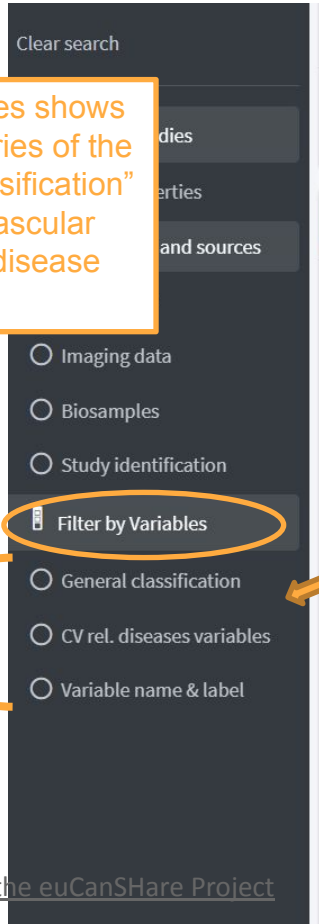


- Variable name & label
  - Search by matching the text in the variable name or label text
  - Using this search may miss some or all relevant variables of the studies, due to e.g.
    - language or abbreviations used, or
    - insufficient information given in the variable label (the information can be written in the broader description of the variable, and that is not used in the search).
  - This search option can be used in the situations when the variables of the study are already familiar to the user.

Functionalities of the Search Tool (10/22)
2) Adding search criteria in variable-level - Filter by variables

E.g. next slides shows some categories of the "General classification" and "Cardiovascular (CV) related disease variables"

Use search criteria for variable-level properties



Query

Individual × AND Cohort × AND General population × AND Biomarkers | Imaging ×

Lists view | Comparison table

Variables 77 041 | Datasets 161 | Studies 26

All | Individual | Harmonization

Show 100 entries

NOTE that studies must have added the variable descriptions to the catalogue and the variables have to be classified into the specific categories (listed in these links), so that it is possible to find that study and its variables by using these variable classes.

Next

Acronym	Name	Type	Study design	Data types, sources, and samples available										Participants	Individual	
															Datasets	Variables
ATBC	ATBC Study	Individual	Cohort	✓	✓	✓	✓	-	-	✓	-	✓	✓	29000	4	279
Brianza	Brianza MONICA Study	Individual	Cohort	✓	✓	✓	✓	-	-	✓	-	✓	✓	4932	6	751
Caerphilly	Caerphilly Prospective Study	Individual	Cohort	✓	✓	✓	✓	✓	-	✓	-	✓	✓	2171	-	-

Functionalities of the Search Tool (11/22)
2) Adding search criteria in variable-level - General classification (part of it)

Here selected are "tobacco" from "lifestyle and behaviours" and "diseases of the circulatory system (I00-I9)" from "diseases"

Click:

Display results

Socio-demographic and economic characteristics — Select All

Information about socio-demographic and economic characteristics.

- ☐ Age/birthdate
- ☐ Marital/partner status
- ☐ Residence
- ☐ Sex/gender
- ☐ Family and household structure
- ☐ Birthplace
- ☐ Language
- ☐ Other socio-demographic and economic characteristics
- ☐ Twin
- ☐ Education
- ☐ Citizenship and immigrant status
- ☐ Labour force and retirement

▾ More

Lifestyle and behaviours — Select All Clear Selection

Information about past and current lifestyle, behaviours and activities.

- ☑ Tobacco
- ☐ Nutrition
- ☐ Transportation
- ☐ Sexual behaviours and orientation
- ☐ Misbehaviour and criminality
- ☐ Alcohol
- ☐ Breastfeeding
- ☐ Personal hygiene
- ☐ Leisure activities
- ☐ Other and unspecified lifestyle information
- ☐ Drugs
- ☐ Physical activity
- ☐ Sleep
- ☐ Technological devices

▾ More

Diseases — Select All Clear Selection

Information about past and current disease as categorized in ICD-10.

- ☐ Certain infectious and parasitic diseases (A00-B99)
- ☐ Neoplasms (C00-D48)
- ☐ Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89)
- ☐ Endocrine, nutritional and metabolic diseases (E00-E90)
- ☐ Mental and behavioural disorders (F00-F99)
- ☐ Diseases of the nervous system (G00-G99)
- ☐ Diseases of the eye and adnexa (H00-H59)
- ☐ Diseases of the ear and mastoid process (H60-H95)
- ☑ Diseases of the circulatory system (I00-I99)
- ☐ Diseases of the

reproductive health history — Select All

irth and current or past reproductive health history of

- ☐ Contraception
- ☐ Pregnancy, delivery and birth
- ☐ Other reproductive health-related information

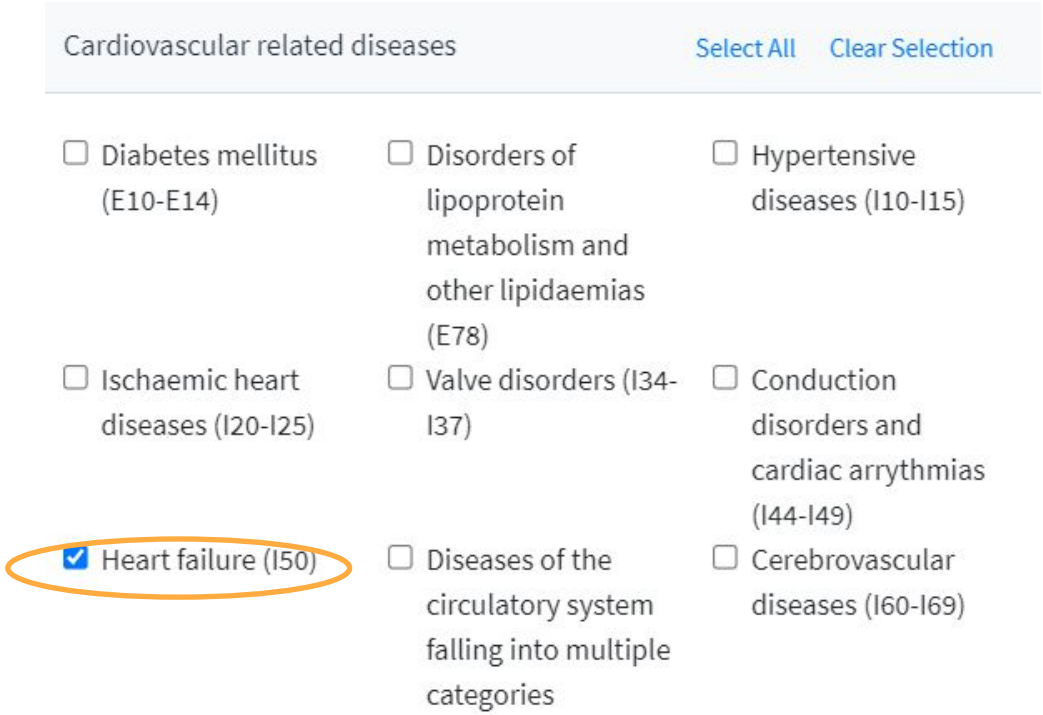
Perception of health, quality of life, development and functional limitations — Select All

Information about perception of general health, quality of life, child development and decline of functional capacities.

- ☐ Perception of health
- ☐ Quality of life
- ☐ Life course development
- ☐ Functional limitations
- ☐ Use of assistive devices
- ☐ Other perception of health, quality of life and functional limitation-related information

Functionalities of the Search Tool (12/22)
2) Adding search criteria in variable-level - Cardiovascular related disease variables

Here selected is "heart failure (I50)" from "cardiovascular related diseases"



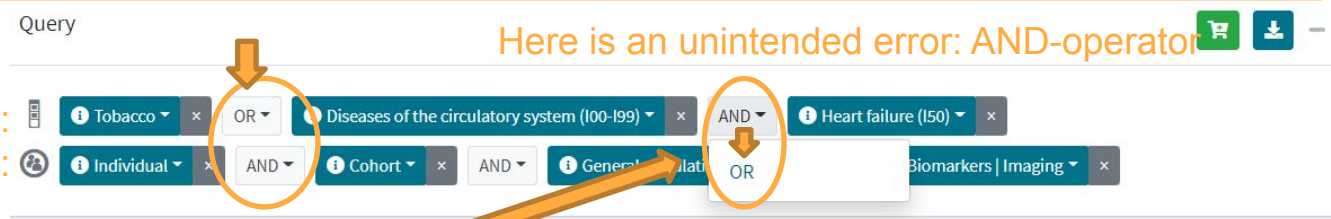
Click: Display results

Functionalities of the Search Tool (13/22)
Results by studies when some study- and variable-level properties selected

At first, check the query used!

Criteria used in variable-level:
Criteria used in study-level:

Here is an unintended error: AND-operator



AND / OR selection:
NOTE that a variable is usually classified into one category only, so AND-operator may not be sensible for filtering by variables.

When adding a criteria from "cardiovascular related disease variables", AND is used mistakenly as the default operator. Change it to OR.

"Heart failure" happen to be a subcategory for the "diseases of the circulatory system", so the result set here is not totally empty, even if AND-operator is used.

Lists view | Comparison table

Variables 278 | Datasets 59 | Studies 11

All | Individual | Harmonization

Show 100 entries

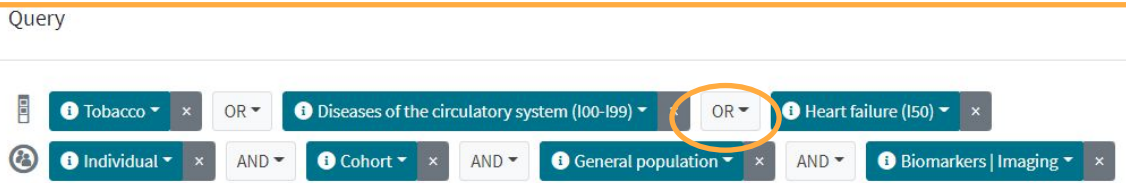
Previous 1 Next

Acronym	Name	Type	Study design	Data types, sources, and samples available										Individual		
														Participants	Datasets	Variables
Catalonia	Catalonia MONICA Study	Individual	Cohort	✓	✓	✓	-	-	-	✓	-	✓	✓	5505	3	20
DAN-MONICA	DAN-MONICA Study	Individual	Cohort	✓	✓	✓	-	-	-	✓	-	✓	✓	7582	8	43

Link to this search

Functionalities of the Search Tool (14/22)
Results by studies when some study- and variable-level properties selected

Corrected query:
OR-operator is used for all variables categories



Lists view | Comparison table

Variables 8 482 | Datasets 147 | Studies 23

All | Individual | Harmonization

Show 100 entries

Previous 1 Next

23 studies include 8482 variables on some of the selected variable classes.

Note that the variables of some studies were not yet (in October 2022) classified to the cardiovascular related disease classes.

Acronym	Name	Type	Study design	Data types, sources, and samples available										Participants	Individual	
															Datasets	Variables
ATBC	ATBC Study	Individual	Cohort	✓	✓	✓	✓	-	-	✓	-	✓	✓	29000	4	148
Brianza	Brianza MONICA Study	Individual	Cohort	✓	✓	✓	✓	-	-	✓	-	✓	✓	4932	6	348
Catalonia	Catalonia MONICA Study	Individual	Cohort	✓	✓	✓	-	-	-	✓	-	✓	✓	5505	4	282

Link to this search

Functionalities of the Search Tool (15/22)
Results by variables when some study- and variable-level properties selected

Lists view | Comparison table

Variables 8 482 | Datasets 147 | Studies 23

List view of variables selected

All | Individual | Harmonization

Show 20 entries

Variables in the specific classifications

Previous 1 2 3 4 5 … 425 Next

	Name	Label	Value type	Annotations	Type	Study	Population	Data Collection Event	Dataset
☐	CIGS	MORGAM variable: "Do you smoke cigarettes now?"	Integer	Tobacco	Collected	KORA	KORA (Augsburg) Study Cohort 01 (Survey S1)	KORA S1 (Augsburg) Study Cohort 01 Baseline	KORA0101_MORGAM
☐	NUMCIGS	MORGAM variable: "On average how many cigarettes do you now smoke a day?" "	Integer	Tobacco	Collected	KORA	KORA (Augsburg) Study Cohort 01 (Survey S1)	KORA S1 (Augsburg) Study Cohort 01 Baseline	KORA0101_MORGAM
☐	EVERCIG	MORGAM variable: "Did you ever smoke cigarettes regularly in the past?"	Integer	Tobacco	Collected	KORA	KORA (Augsburg) Study Cohort 01 (Survey S1)	KORA S1 (Augsburg) Study Cohort 01 Baseline	KORA0101_MORGAM

Link to this search

Functionalities of the Search Tool (16/22)
Results as comparison table of variable classes, by the studies

Some variable-level criteria need to be selected to see the comparison table.

Lists view | Comparison table

Comparison table by studies selected

To see the time when the data are collected: Data collection events

Study | Dataset

Data Collection Event | Filter

All | Individual | Harmonization

	Cardiovascular related diseases ×	Lifestyle and behaviours ×	Diseases ×
Study	Heart failure (I50) ×	Tobacco ×	Diseases of the circulatory system (I00-I99) ×
	278	1 213	7 269
KORA	10	67	313
Brianza	0	42	306
Catalonia	20	30	252
DAN-MONICA	43	66	455
ESTHER	2	12	108
Estonia	3	15	117

Null cell

Click the number of variables to see the list of variables in this class for the study.

23 studies listed

Link to this search

Functionalities of the Search Tool (17/22)
Results as comparison table of variable classes, by the data collection events

Lists view | Comparison table

Data collection event selected

Study | Dataset

All | Individual | Harmonization

Data Collection Event | Filter

To filter the search results to include those rows that includes variables in all classes, i.e. remove those having null cell ("full coverage").

Population/Data Collection Event (DCE)			Cardiovascular related diseases ×	Lifestyle and behaviours ×	Diseases ×
Study	**Population**	**Data Collection Event**	**Heart failure (I50) ×**	**Tobacco ×**	**Diseases of the circulatory system (I00-I99) ×**
			278	**1 213**	**7 269**
KORA	KORA (Augsburg) Study Cohort 01 (Survey S1)	KORA S1 (Augsburg) Study Cohort 01 Baseline 1984-10 to 1985-05	2	16	17
		KORA S1 (Augsburg) Study Cohort 01 mortality and disease outcome follow-up 1984-10 to 2009-12	0	0	27
	KORA (Augsburg) Study Cohort 02 (Survey S2)	KORA S2 (Augsburg) Study Cohort 02 Baseline 1989-10 to 1990-06	2	15	17

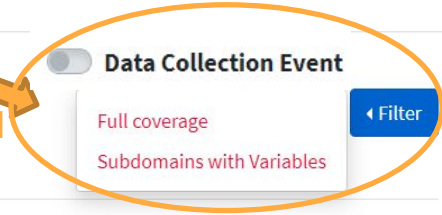
Null cells

Link to this search

Functionalities of the Search Tool (18/22)
Results as full coverage comparison table, by the studies

Lists view | Comparison table

Filtering “full coverage” selected
and
“data collection event” unselected



Study | Dataset

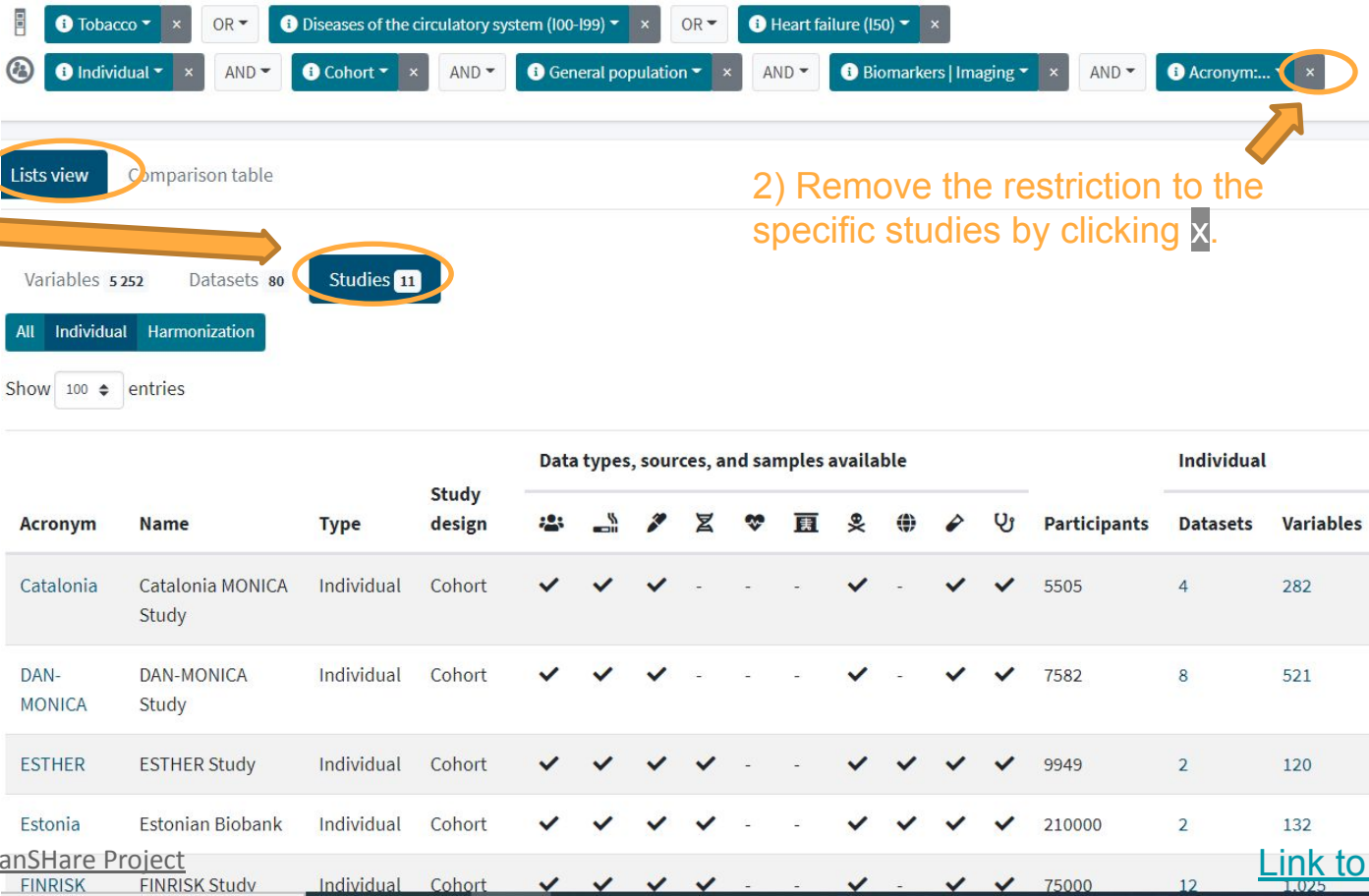
All | Individual | Harmonization

	Cardiovascular related diseases ×	Lifestyle and behaviours ×	Diseases ×
Study	Heart failure (I50) ×	Tobacco ×	Diseases of the circulatory system (I00-I99) ×
	278	758	4 494
KORA	10	67	313
Catalonia	20	30	252
DAN-MONICA	43	66	455
ESTHER	2	12	108
Estonia	3	15	117
FINRISK	77	96	929

Null cells removed, 11 studies listed

Link to this search

Functionalities of the Search Tool (19/22)
Removing the full coverage restrictions, results by studies



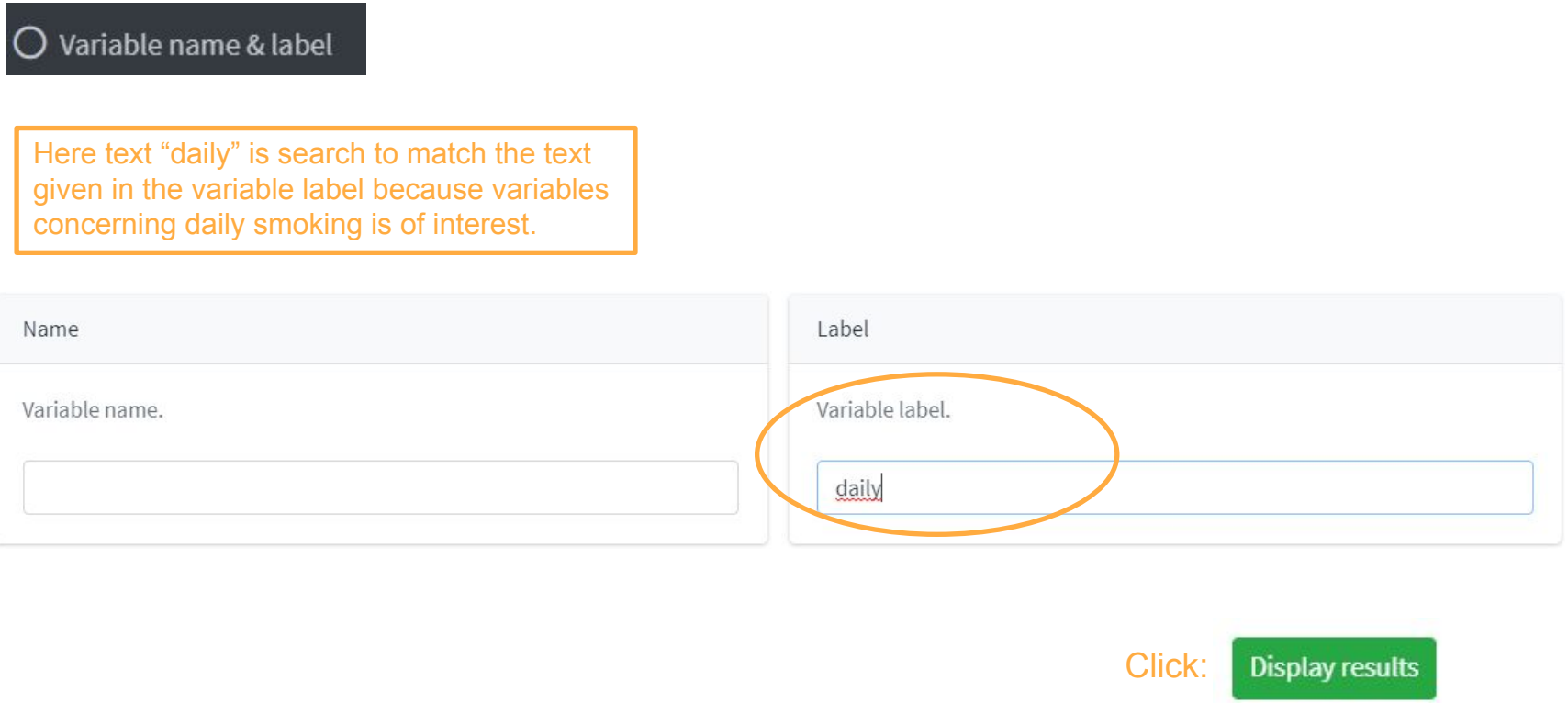
1) Select list view and results by studies.

2) Remove the restriction to the specific studies by clicking x.

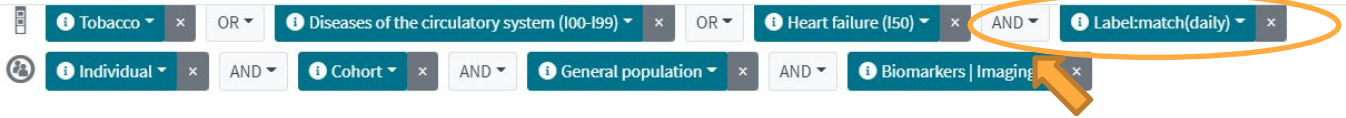
Acronym	Name	Type	Study design	Data types, sources, and samples available	Participants	Individual		Harmonization	
						Datasets	Variables	Datasets	Variables
Catalonia	Catalonia MONICA Study	Individual	Cohort	✓ ✓ ✓ - - - ✓ - ✓ ✓	5505	4	282	-	-
DAN-MONICA	DAN-MONICA Study	Individual	Cohort	✓ ✓ ✓ - - - ✓ - ✓ ✓	7582	8	521	-	-
ESTHER	ESTHER Study	Individual	Cohort	✓ ✓ ✓ ✓ - - ✓ ✓ ✓ ✓	9949	2	120	-	-
Estonia	Estonian Biobank	Individual	Cohort	✓ ✓ ✓ ✓ - - ✓ ✓ ✓ ✓	210000	2	132	-	-
FINRISK	FINRISK Study	Individual	Cohort	✓ ✓ ✓ ✓ - - ✓ - ✓ ✓	75000	12	1,025	-	-

Link to this search

Functionalities of the Search Tool (20/22)
Adding search criteria in variable-level - Searching text in the variable label



Functionalities of the Search Tool (21/22)
Results by variables when also matching text with variable label



AND-operator should be used here, as daily smoking variables are of interest (i.e. categorized to tobacco AND label contains “daily”).

Lists view | Comparison table

List view of variables selected

Variables 287 | Datasets 80 | Studies 23

All | Individual | Harmonization

Show 100 entries

Previous 1 2 3 N

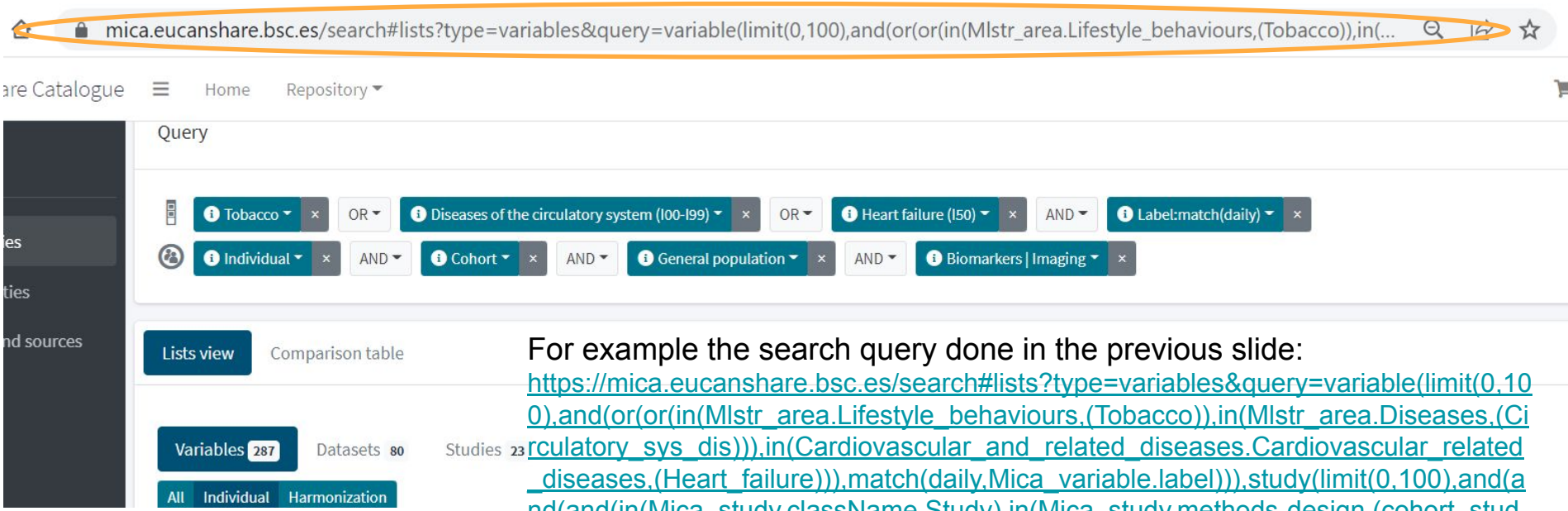
Name	Label	Value type	Annotations	Type	Study	Population	Data Collection Event	Dataset
MAXCIGS	MORGAM variable: "What is the highest average daily number of cigarettes you have ever smoked for as long as a year?"	Integer	Tobacco	Collected	KORA	KORA (Augsburg) Study Cohort 01 (Survey S1)	KORA S1 (Augsburg) Study Cohort 01 Baseline	KORA0101_MORGAM
DSMOKER	MORGAM variable: Daily cigarette smoker	Integer	Tobacco	Collected	KORA	KORA (Augsburg) Study Cohort 01 (Survey S1)	KORA S1 (Augsburg) Study Cohort 01 Baseline	KORA0101_MORGAM
STOPAGE	MORGAM variable: Age when the person stopped smoking cigarettes daily	Integer	Tobacco	Collected	KORA	KORA (Augsburg) Study Cohort 01 (Survey S1)	KORA S1 (Augsburg) Study Cohort 01 Baseline	KORA0101_MORGAM

“MAXCIGS” is shown later as an example of the variable description.

Link to this search

Functionalities of the Search Tool (22/22)
Saving the search query for later use

Copy-paste the URL in the address bar and save the URL for later use.
That link can be used to do the the exactly same search again.



For example the search query done in the previous slide:
https://mica.eucanshare.bsc.es/search#lists?type=variables&query=variable(limit(0,100),and(or(or(in(Mlstr_area.Lifestyle_behaviours,(Tobacco)),in(Mlstr_area.Diseases,(Circulatory_sys_dis))),in(Cardiovascular_and_related_diseases.Cardiovascular_related_diseases,(Heart_failure))),match(daily,Mica_variable.label))),study(limit(0,100),and(and(and(in(Mica_study.className,Study),in(Mica_study.methods-design,(cohort_study))),in(Mica_study.populations-recruitment-dataSources,(general_population))),in(Mica_study.populations-dataCollectionEvents-dataTypes,(biomarkers,imaging))))

Finding the harmonization potential across the studies

- The variable descriptions from different studies can be compared in the catalogue to find the possible definition of the harmonized variable.
- The variable descriptions of the catalogue includes
  - Variable definition, type and unit
  - Category values and labels for the categorical variables
  - In some cases, the number of available and missing values

Finding the harmonization potential (1/5)
Viewing the variable description of the specific data collection event (1/2)

Here presented is the variable description of the “MAXCIGS” collected in the baseline of the KORA S1 Survey.

MAXCIGS Variable name

MORGAM variable: "What is the highest average daily number of cigarettes you have ever smoked for as long as a year?" Variable label

"See the specific description of the variable in the MORGAM website. Possible broader description of the variable

Overview

Value type	Integer
Nature	Continuous
Entity type	Participant

Categories Possible category names and labels (MAXCIGS is continuous variable, and only missing categories are defined)

Name	Label	Missing
888	irrelevant if EVERCIG = 2	✔
999	insufficient data	✔

Definition

Dataset	KORA0101_MORGAM
Study	KORA
Population	KORA (Augsburg) Study Cohort 01 (Survey S1)
Data Collection Event	KORA S1 (Augsburg) Study Cohort 01 Baseline

Classes in which the variable is classified i.e. annotations

Annotations

Source	Questionnaire
Target	Participant
Lifestyle and behaviours	Tobacco

Annotations: “Questionnaire” = information was collected in a questionnaire, “Participant” = information is about the study participant, “Tobacco” = information is about the consumption tobacco of in any form

Finding the harmonization potential (2/5)
Viewing the variable description of the specific data collection in the study (2/2)

Summary statistics of the variable show the availability information for the variable in the specific dataset (here MAXCIGS in KORA S1 Baseline), if the dummy data on the missingness are provided.

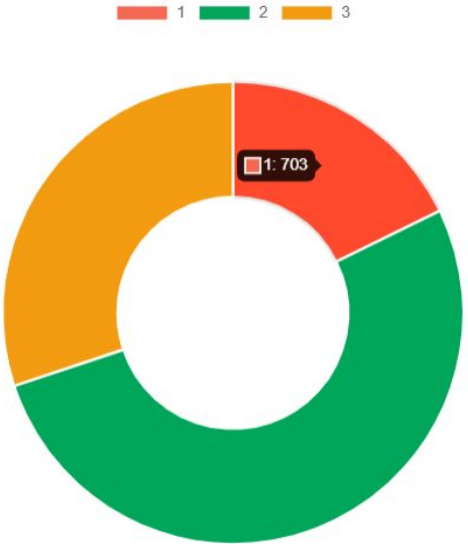
Summary Statistics

N	3,969
N with values	2,774
N missings	1,195

Number of observations in the dataset

Number of missing values

Proportions of the missing categories as pie chart



"MAXCIGS" variable is relevant only for the current or ex-smokers, and "replaceable to valid" for ever non-smokers.

Missing categories defined for the study and the frequencies for the available and missing values

Frequencies

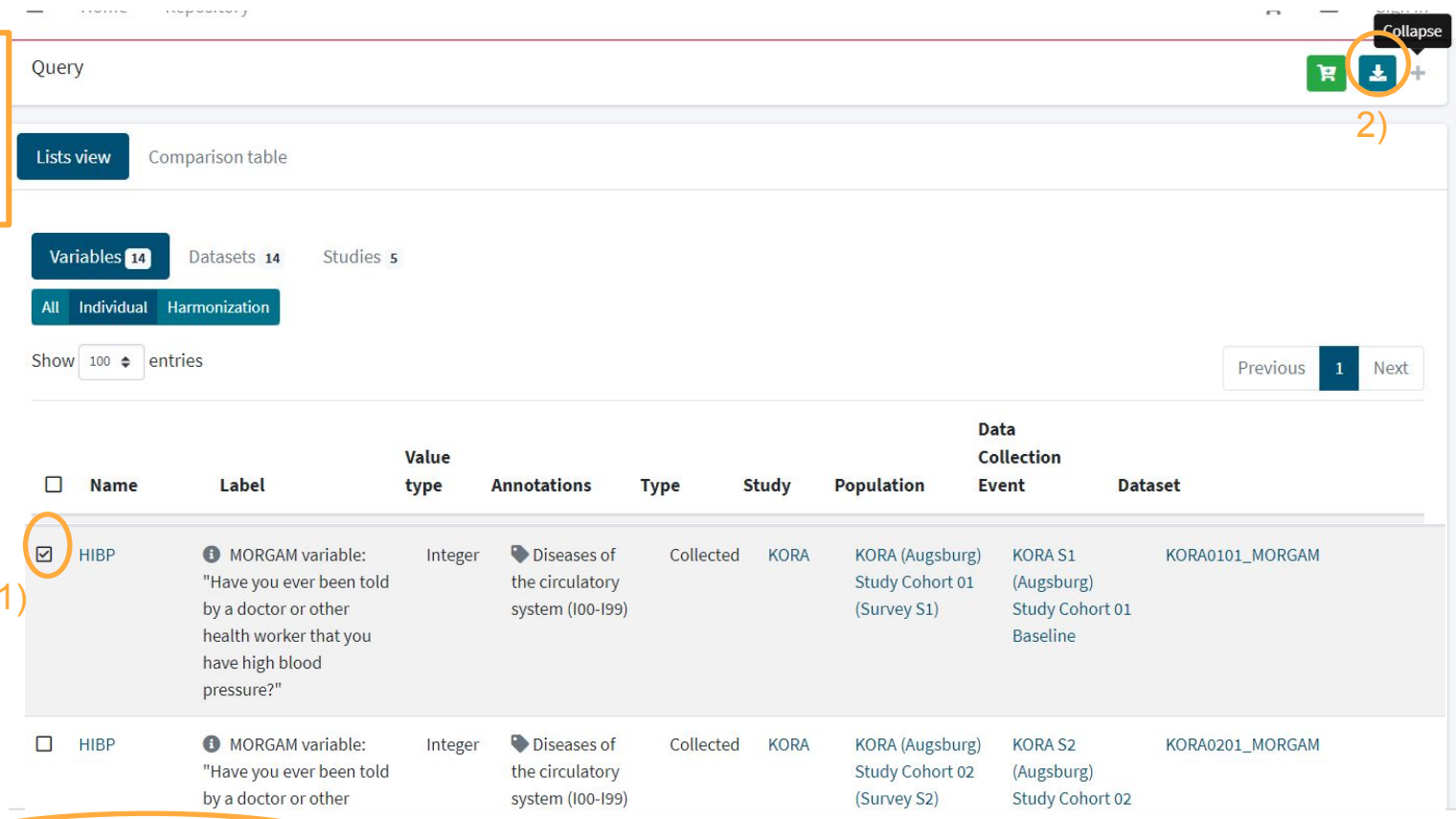
Value	Frequency	% with values	% missings	%	Missing
1 Valid value	703	25.34		17.71	
2 Missing, but replaceable to valid	2,071	74.66		52.18	
3 Missing (not by design)	1,195		100	30.11	✓
4 Missing by design	0		0	0	✓
5 Not applicable	0		0	0	✓
N/A Empty values	0		0	0	✓

Finding the harmonization potential across the studies

- On next slides, a simple example set of variables shows the ways that data catalogue can be used for finding:
  - harmonization potential across the studies/cohorts
  - common definition for the harmonized variable
- Small number of similar variables on high blood pressure diagnoses across studies are selected (selected from this list):
  - “HIBP” of KORA S1 and ESTHER Baselines (harmonized MORGAM variable)
  - “hcbphigh” of CAHHM CPTP cohort Enrollment
  - “s0_hypert” of SHIP-START-0
  - variables “6150_0_0”, “6150_0_1”, “6150_0_2”, and “6150_0_3” of UK Biobank Recruitment

Finding the harmonization potential (3/5)
Downloading the results as a CSV file

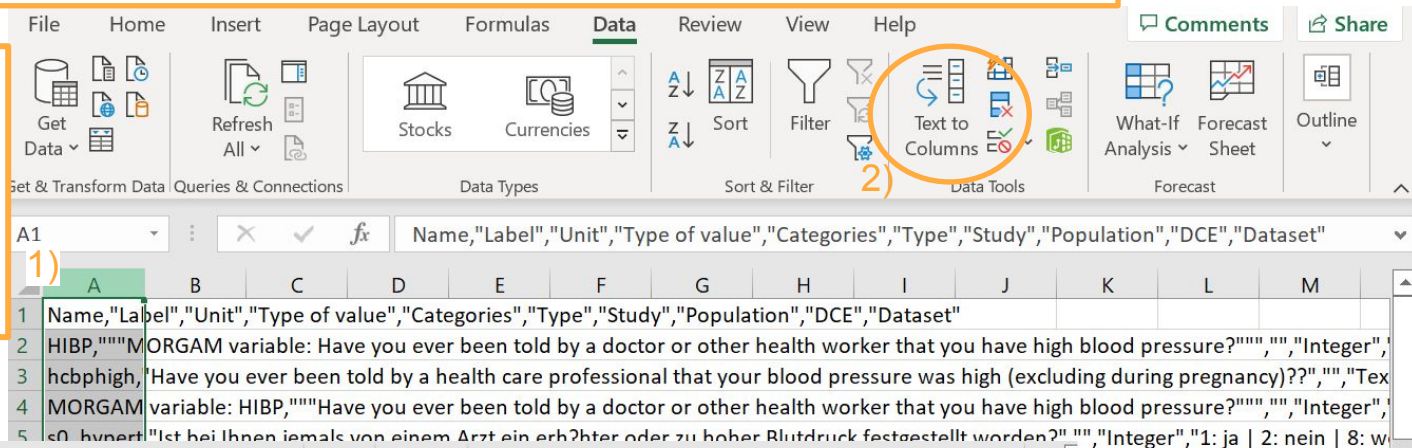
1) Tick the desired variables
2) Click “Download”
3) Open CSV



CSV file (comma ”,” separated table of the search results) is downloaded. The file can be opened using Excel.

Finding the harmonization potential (4/5)
Opening the results and comparing variables

1) In Excel, select column A
2) In **"Data"** tab click **"Text To Colums"** and follow the converting steps (delimeted by comma ",") ->
3) Splitted to the columns.



3)

	A	B	C	D	E
1	Name	Label	Unit	Type of	Categories
2	HIBP	MORGAM variable: "Have you ever been told by a doctor or other health worker that you have high blood pressure?"		Integer	1: yes | 2: no | 9: insufficient data
3	hcbphigh	Have you ever been told by a health care professional that your blood pressure was high (excluding during pregnancy)?		Textual	
4	HIBP	MORGAM variable: "Have you ever been told by a doctor or other health worker that you have high blood pressure?"		Integer	1: yes | 2: no | 9: insufficient data
5	s0_hypert	Have you ever been diagnosed with high blood pressure by a doctor?		Intege	1: yes | 2: no | 8: don't know | 9: refused to a
6	6150_0_0	Touchscreen Health and medical history Medical conditions: Vascular heart problems diagnosed by doctor		Textual	1: Heart attack | 2: Angina | 3: Stroke | 4: High blood pressure | -7: None of the above | -3: Prefer not to answer
7					

All variables 6150_0_0 - 6150_0_3 have the same description: ACE touchscreen question: Has a doctor ever told you that you have had any of the following conditions? (You can select more than one answer).

Note that using **"Text To Colums"** for Categories delimited by "|" (copy first column E to the most right of the table), classes can be separated further into different cells.

Finding the harmonization potential (5/5)
Comparing variables across studies

The definition of this CAHMM variable needs more investigation as the category values are not given in the catalogue. Here it is assumed that yes & no categories are included.

The definition of the UKBB variable needs more investigation, but here it is assumed that this 4th variable indicates the answers to the "4 - High blood pressure" option in the questionnaire.

	A	B	C	D	K	L	M	N
1	**Name**	**Label**	**Unit**	**Type of**	**Categories**			
2	HIBP	MORGAM variable: "Have you ever been told by a doctor or other health worker that you have high blood pressure?"		Integer	1: yes	2: no	9: insufficient data	
3	hcbphigh	Have you ever been told by a health care professional that your blood pressure was high (excluding during pregnancy)?		Textual	?: yes?	?: no?		
4	HIBP	MORGAM variable: "Have you ever been told by a doctor or other health worker that you have high blood pressure?"		Integer	1: yes	2: no	9: insufficient data	
5	s0_hypert	Have you ever been diagnosed with high blood pressure by a doctor?		Integer	1: yes	2: no	8: Don't know	9: Refused
6	6150_0_3	Touchscreen Health and medical history Medical conditions: 4: high blood pressure diagnosed by doctor		Textual	?: yes?	?: no?		-3: Prefer not to answer

"Has a doctor ever told you that you have had any of the following conditions?" (You can select more than one answer).

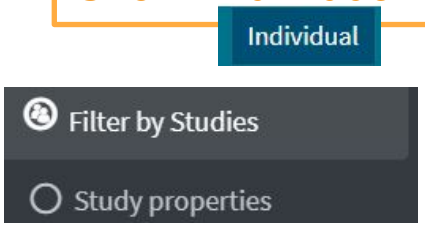
A possible definition for a harmonized variable:
"Have you ever been told by a doctor or other health professional that you have high blood pressure?"
1: yes
2: no
9: missing

Example search

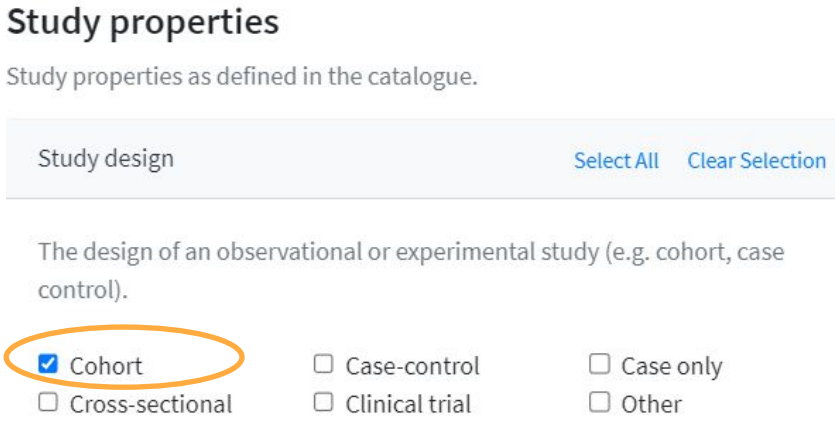
Question: 1) Which cohort studies have collected variables on all these categories: cerebrovascular diseases, alcohol, medication/supplement intake, and quality of life?
2) What are the names of the quality of life variables in these studies?

Solution: Filtering by both studies and variables is needed and the results must be shown as the comparison table for further filtering off the null cells of the variable categories.

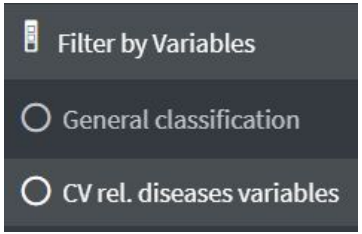
Solution for example, step 1/5 – Start a new search: https://mica.eucanshare.bsc.es/search#
Click ”Individual” and go to…



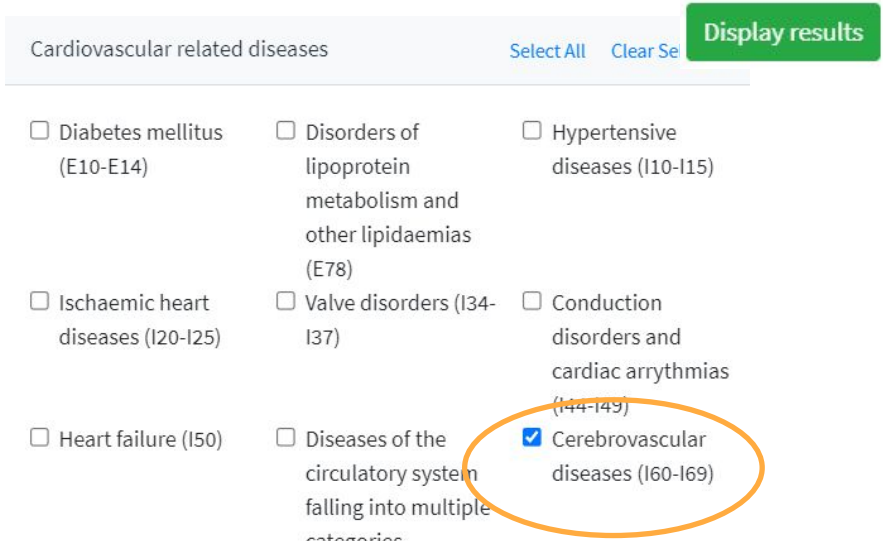
…Study properties
Select “cohort”



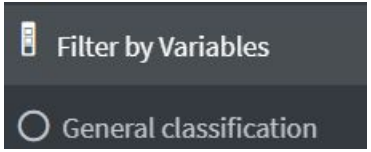
Click ”Display results” and go to…



…Cardiovascular related disease variables
Select “cerebrovascular diseases (I60-I69)”



Solution for example, step 2/5 - Go to ”general classification”



Find and select “alcohol”, “medication and supplement intake” and “quality of life”

Lifestyle and behaviours Select All Clear Selection

Information about past and current lifestyle, behaviours and activities.

- ☐ Tobacco
- ☐ Nutrition
- ☐ Transportation
- ☐ Sexual behaviours and orientation
- ☐ Misbehaviour and criminality
- ☑ Alcohol
- ☐ Breastfeeding
- ☐ Personal hygiene
- ☐ Leisure activities
- ☐ Other and unspecified information
- ☐ Drugs
- ☐ Physical activity
- ☐ Sleep
- ☐ Technological

Medication and supplements Select All Clear Selection

Information about medication (whether prescribed or over the counter), including drugs and supplements (e.g. vitamins, plant extracts) used to treat or prevent diseases or to alleviate symptoms of diseases.

- ☑ Medication and supplement intake
- ☐ Posology and protocol of administration
- ☐ Other and unspecified pharmacological interventions

▾ More

Perception of health, quality of life, development and functional limitations Select All Clear Selection

Information about perception of general health, quality of life, child development and decline of functional capacities.

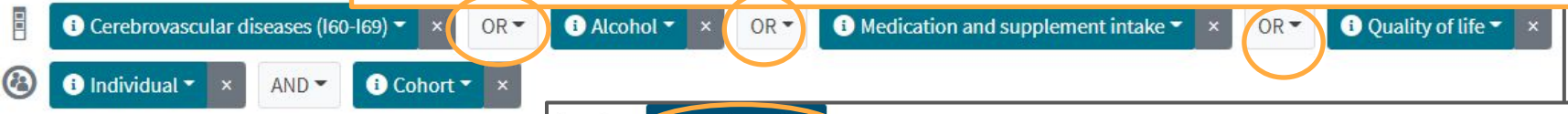
- ☐ Perception of health
- ☑ Quality of life
- ☐ Life course development
- ☐ Functional limitations
- ☐ Use of assistive devices
- ☐ Other perception of health, quality of life and functional limitation-related information

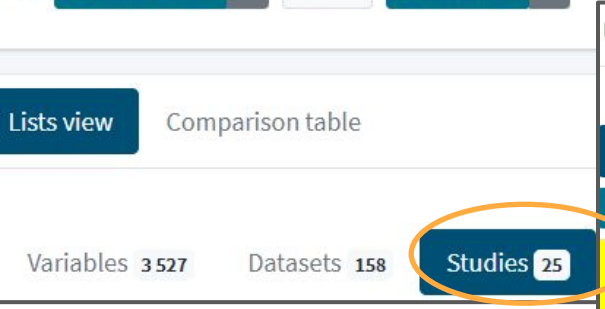
Then click:

Display results

Solution for example, step 3/5 - Check the query

OR-operator is needed here as 1 variable can't be classified to all these classes at the same time.





25 studies have variables in some of these classes. Then select comparison table to further filtering off the studies that don't have all these variable classes collected.

Null cells

Lists view | Comparison table

Select Filter: "full coverage"

Study | Dataset

All | Individual | Harmonization

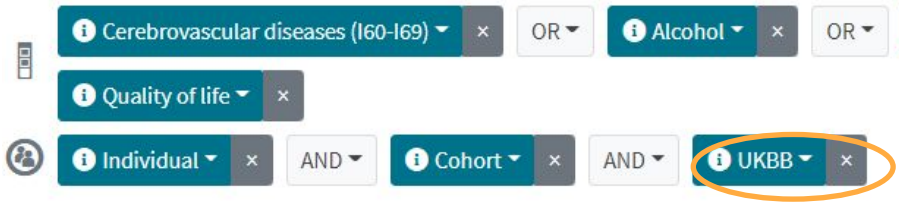
Filter

Full coverage

Subdomains with Variables

Study	Cardiovascular related diseases × / Cerebrovascular diseases (I60-I69) ×	Lifestyle and behaviours × / Alcohol ×	Perception of health, quality of life, development and functional limitations × / Quality of life ×	Medication and supplements × / Medication and supplement intake ×
	1 608	531	19	1 369
KORA	62	20	0	15
Brianza	81	12	0	14
CAHHM	0	148	12	290
Catalonia	54	2	0	11

Solution for example, step 4/5 - See the full coverage results



Lists view | Comparison table

1 study, UKBB, has variables in all these classes.

Study | Dataset

Data Collection Event | Filter

All | Individual | Harmonization

Study	Cardiovascular related diseases ×	Lifestyle and behaviours ×	Perception of health, quality of life, development and functional limitations ×	Medication and supplements ×
	Cerebrovascular diseases (I60-I69) ×	Alcohol ×	Quality of life ×	Medication and supplement intake ×
	20	109	7	677
UKBB	20	109	7	677

Click to see the list of quality of life variables in the UKBB study

Solution for example, step 5/5 - See the list of variables

Lists view | Comparison table

Variables 7 | Datasets 3 | Studies 1

All | Individual | Harmonization

7 quality of life variables

Show 20 entries

Previous 1 Next

Name	Label	Value type	Annotations	Type	Study	Population	Data Collection Event	Dataset
26413_0_0	Baseline characteristics Indices of Multiple Deprivation: Health score (England)	Decimal	Quality of life	Collected	UKBB	UK Biobank	UK Biobank - Recruitment	UKBB_Baseline_assessment
26420_0_0	Baseline characteristics Indices of Multiple Deprivation: Health score (Wales)	Decimal	Quality of life	Collected	UKBB	UK Biobank	UK Biobank - Recruitment	UKBB_Baseline_assessment
26430_0_0	Baseline characteristics Indices of Multiple Deprivation: Health score (Scotland)	Decimal	Quality of life	Collected	UKBB	UK Biobank	UK Biobank - Recruitment	UKBB_Baseline_assessment
20458_1_0	Mental health Happiness and subjective wellbeing: General happiness	Text	Quality of life	Collected	UKBB	UK Biobank	UK Biobank - Mental health questionnaire	UKBB_Mental_health_enhancement_data
20459_1_0	Mental health Happiness and subjective wellbeing: General	Text	Quality of life	Collected	UKBB	UK Biobank	UK Biobank - Mental health questionnaire	UKBB_Mental_health_enhancement_data

All rows are not presented here

Answers to the example search

1) Which cohort studies have collected variables on all these categories: cerebrovascular diseases, alcohol, medication/supplement intake, and quality of life?

Answer: UK Biobank (UKBB)

Link to this answer

2) What are the names of the quality of life variables in these studies?

Answer: UKBB variable names are 26413_0_0, 26420_0_0, 26430_0_0, 20458_1_0, 20459_1_0, 120070_1_0, and 120097_1_0.

Link to this answer