

Instructions for Setting Up Cohorts and Creating Dataframes in AoU CohortBuilder and DatasetBuilder

Introduction: This guide provides detailed steps on how to set up specific cohorts and create data frames using AoU CohortBuilder and DatasetBuilder. By following these instructions, you'll be prepared to use the Condition Association code template designed to explore potential associations between any two conditions defined by you—the researcher.

Step 1: Define Your Conditions

- **Condition 1:** This will be the primary condition you're interested in.
- **Condition 2:** Another condition with which you seek to explore a potential association with condition 1.

Step 2: Set Up Cohorts

- **Case Cohort:** A cohort that includes people with condition 1.
- **Control Cohort:** A cohort that excludes people with condition 1.
- **Condition Cohort:** A cohort that includes people with condition 2.

Step 3: Create Data Frames

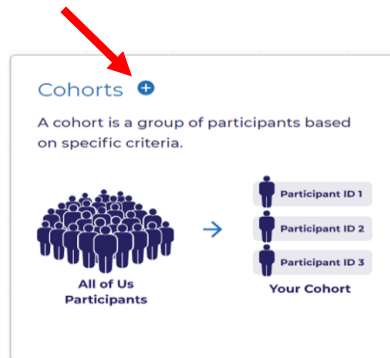
- For the case and control cohorts, you'll need data frames with the following variables:
- **From the DatasetBuilder:**
 1. `person_id`
 2. `date_of_birth`
 3. `race`
 4. `ethnicity`
 5. `sex_at_birth`
- **Created in the Code Template Notebook:**
 1. `predictor`
 2. `outcome`
 3. `age`

Conclusion: Once you've set up the cohorts and created the necessary data frames, you can proceed to utilize the code template to analyze and understand the potential association between the conditions you've defined.

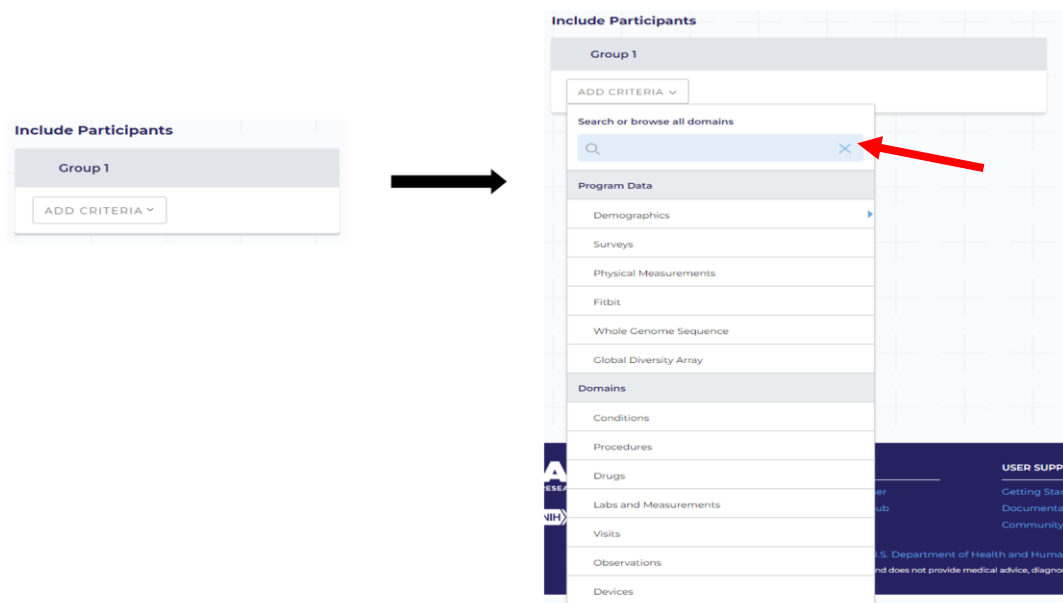
Define Conditions & Set Up Cohorts

Creating the case cohort:

1. Go to the CohortBuilder in the All of Us Researcher Workbench.



2. Under **Group 1**, search for **condition 1**.

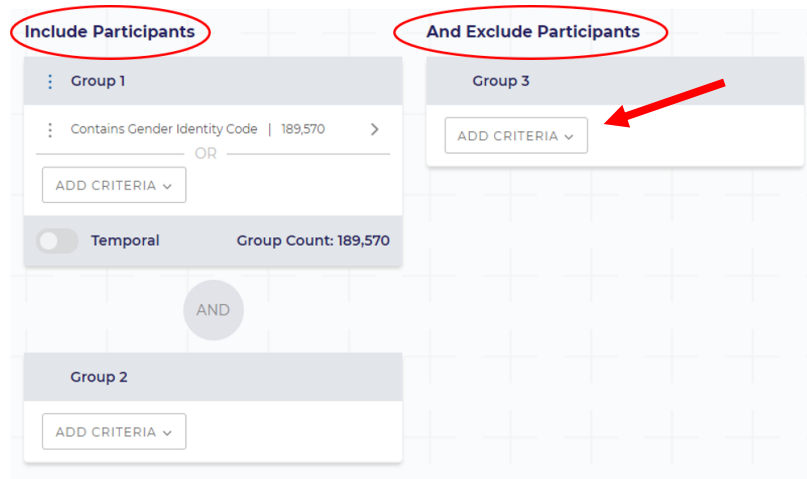


3. Add **condition 1** to the inclusion criteria.

The image shows the 'Conditions' search results page. At the top, there is a search bar with 'F33' entered. Below the search bar, there is a message: 'There are 19,490 participants for the standard version of the code you searched. Return to source code. There are 11 results in All of Us Dataset v5.' Below this message is a table with the following columns: Name, Code, Vocab, Source/Standard, Concept Id, Roll-up Count, Item Count, and View Hierarchy. The first row of the table is highlighted with a red arrow pointing to the 'Name' column.

Name	Code	Vocab	Source/Standard	Concept Id	Roll-up Count	Item Count	View Hierarchy
Recurrent major depression	66344007	SNOMED	Standard	4282316	19,490	7,476	

4. Go to exclusion criteria under **Group 3**, then select **Demographics**.
 - a. For the variables **race**, **ethnicity**, and **sex_at_birth**, select the values you **do not** want included in the analysis (for example, you can exclude those who answered “Skip”).



5. Name and save the cohort.



Creating the control cohort:

1. Go to the CohortBuilder in the All of Us Researcher Workbench.
2. Under **Group 1**, select **Demographics**, then select **Age**.
 - a. This will be a dummy inclusion criteria, so we can access the exclusion criteria.
 - b. Keep the age range as it is and save the selection.
 - c. The exclusion criteria should now be available.
3. Under **Group 3**, add **condition 1** to the exclusion criteria.
4. Go to exclusion criteria under **Group 4**, then select **Demographics**.
 - a. For the variables **race**, **ethnicity**, and **sex_at_birth**, select the values you **do not** want included in the analysis (for example, you can exclude those who answered “Skip”).
 - b. DO NOT put the exclusion criteria for all variables under one group, or it will be interpreted as an ‘or’ statement.
5. Name and save the cohort

Creating the condition cohort:

1. Go to the CohortBuilder in the All of Us Researcher Workbench.
2. Under **Group 1**, search for **condition 2**.
3. Add **condition 2** to the inclusion criteria.
4. Go to exclusion criteria under **Group 3**, then select **Demographics**.
 - a. For the variables **race**, **ethnicity**, and **sex_at_birth**, select the values you **do not** want included in the analysis (for example, you can exclude those who answered “Skip”).
5. Name and save the cohort.

Set Up & Export Data Frames

Please note that these data frames will be imported to the bottom of the template. You will have to move the cells up to the top of the notebook so they are the first things to run.

Creating a data frame for the case cohort:

- 1) Under **Select Cohorts**, check the box next to the case cohort you created
- 2) Under **Select Concept Sets**, check the prepackaged **Demographics** concept set
- 3) Under **Select Values (Columns)**, check the boxes for **person id**, **sex at birth**, **date of birth**, **race**, and **ethnicity**. You **must** import these values for the dataset to work in the template.
- 4) Select **Save** and name your data frame
- 5) Select **Export**, and choose the existing notebook “Condition Association Template.ipynb”

Creating a data frame for the control cohort:

- 1) Under **Select Cohorts**, check the box next to the control cohort you created
- 2) Under **Select Concept Sets**, check the prepackaged **Demographics** concept set

- 3) Under **Select Values (Columns)**, check the boxes for **person id**, **sex at birth**, **date of birth**, **race**, and **ethnicity**. You **must** import these values for the dataset to work in the template.
- 4) Select **Save** and name your data frame
- 5) Select **Export**, and choose the existing notebook “Condition Association Template.ipynb”

Creating a data frame for the condition cohort:

- 1) Under **Select Cohorts**, check the box next to the condition cohort you created
- 2) Under **Select Concept Sets**, check the prepackaged **Demographics** concept set
- 3) Under **Select Values (Columns)**, check the boxes for **person id**. You **must** import this value for the dataset to work in the template.
- 4) Select **Save** and name your data frame
- 5) Select **Export**, and choose the existing notebook “Condition Association Template.ipynb”