

All of Us Research Workbench Overview



Introduction

Welcome to the *All of Us* Researcher Workbench!

All workbench analyses happen in the context of a “workspace.” Within a workspace, you can select participants using the Cohort Builder tool. Another tool, the Dataset Builder, allows you to select concepts and values for your cohort. You can then save and export the dataset to a “notebook” for analysis.

For illustration, let’s consider research on “studying correlations between Type 2 diabetes and hypertension” for this quick tour.

Create a New Workspace

A workspace is your place to store and analyze data for a specific project. You can share this workspace with other users allowing them to view or edit your work. The dataset in the workspace is in OMOP common data model format. We have more detailed articles to help you understand OMOP data model in the Researcher Workbench's User Support Hub.

When you create your workspace, you will be prompted to state your research purpose. For example, when you create a workspace to study type 2 diabetes, you could enter the following for research purpose: "I will use this workspace to investigate the correlation between hypertension and type 2 diabetes." You can also find examples for help with filling out your workspace descriptions in the User Support Hub.

Create a new Workspace (Required) ⓘ

Workspace Name

All of Us Dataset v3 ⓘ

All of Us Billing account

The All of Us Program provides \$300 in free credits per user. Please refer to this article to learn more about the free credit program and how it can be used. Once you have used up your free credits, you can request additional credits by contacting support.

Select account

Use All of Us free credits ▾

View free credits balance

Research Use Statement Questions

Best practices for Research Use Statement questions

The All of Us Research Program requires each data user of the All of Us data to provide a meaningful description of the intended purpose of data use for each workspace they create. To provide transparency to All of Us Program participants, your answers below will be made available publicly in the Research Hub Directory on our public website. **Your responses will not be used to make decisions about data access.**

Note that you are required to create separate Workspaces for each project for which you access All of Us data, hence the responses below are expected to be specific to the project for which you are creating this particular Workspace.

1. What is the primary purpose of your project? Publicly displayed ⓘ

☐ Research purpose >

☐ Educational Purpose

The data will be used for education purposes (e.g. for a college research methods course, to educate students on population-based research approaches).

☐ For-Profit Purpose

The data will be used by a for-profit entity for research or product or service development (e.g. for understanding drug responses as part of a pharmaceutical company's drug development or market research efforts).

All of Us
RESEARCH PROGRAM

Build a Cohort

A “cohort” is a group of participants you are interested in researching. The Cohort Builder allows you to create and review cohorts and annotate participants in your study group.

For example, you can build a cohort called “diabetes cases” to include people who have been diagnosed with type II diabetes, using a combination of billing codes and laboratory values. You can also have a “controls” cohort. Once you build your cohorts, you can go through and manually review the records for each participant and decide if you want to include or exclude them from your cohort and make specific annotations/notes to each record.



DATA
ANALYSIS
ABOUT

Group 1

Contains Demographics Code | 526,890

OR

ADD CRITERIA

Program Data
Surveys
Physical Measurements

Domains
Demographics
Conditions
Procedures
Drugs
Measurements
Visits

Group Count: 526,890

AND

Group 3

Contains Drugs Code | 74,755

OR

ADD CRITERIA

Temporal

Group Count: 74,755

AND

Group 4

ADD CRITERIA

Total Count: 485,590

Results by Gender

Gender

Female
Male
Unknown

0 100k 200k 300k 400k 500k 600k

Participants

Results By Gender, Age Range, and Race

Female 19-44
Female 45-64
Female > 65

0% 20% 40% 60% 80% 100%

CREATE COHORT

i

Synthetic data used for training purposes

All of Us
RESEARCH PROGRAM

Build a Concept Set

Concepts describe information in a patient's medical record, such as a condition they have, a prescription they are taking or their physical measurements. In the workbench we refer to subject areas such as conditions, drugs, measurements, etc. as “domains.” You can search for and save collections of concepts from a particular domain as a “concept set.”

For example, if you want to select height, weight, and blood pressure information (concepts) from your “diabetes cases” cohort, you can search for the 3 concepts from the “Measurements” domain and call it “biometrics” concept set. You can then use the Dataset Builder to export that information about your cohort into a notebook.



DATA

ANALYSIS

ABOUT

Search Concepts

☒ Standard concepts only

EHR Domain

Conditions

22668

concepts in this domain.

946237 participants in domain.

Browse Domain

Drug Exposures

3812

concepts in this domain.

892671 participants in domain.

Browse Domain

Measurements

2623

concepts in this domain.

944511 participants in domain.

Browse Domain

Procedures

15879

concepts in this domain.

940058 participants in domain.

Browse Domain

Survey Questions

The Basics

122

survey questions with

567437 participants

Survey includes participant demographic information.

Browse Survey

Overall Health

26

survey questions with

567857 participants

Survey provides information about how participants report levels of individual health.

Browse Survey

Lifestyle

62

survey questions with

568120 participants

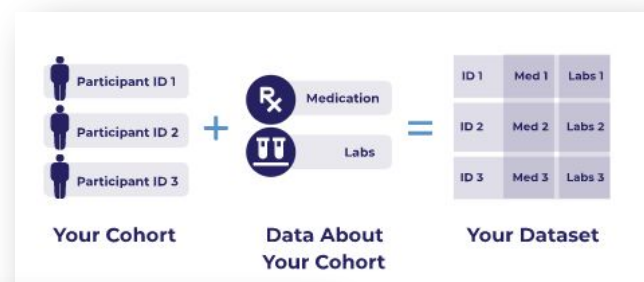
Survey includes information on participant smoking, alcohol and recreational drug use.

Browse Survey

Build a Dataset

Datasets are analysis-ready tables that can be exported to a notebook. You can build and preview a dataset for one or more cohorts by selecting the desired concept sets and values for the cohorts.

For example, select your cohort called “diabetes cases” and your concept set called “biometrics” and then the values from the concept set you want to see. Before exporting the dataset, you can preview the resulting data frame that will be sent to the notebook.



DATA

ANALYSIS

ABOUT

Data Sets

Build a data set by selecting the variables and values for one or more of your cohorts. Then export the completed Data Set to Notebooks where you can perform your analysis

1 Select Cohorts (Participants)

Prepackaged Cohorts

☐ All Participants

Workspace Cohorts

☒ diabetes cases

2 Select Concept Sets (Rows)

☐ Demographics

☐ All Surveys

Workspace Concept Sets

☒ Biometrics

☐ test 1

3 Select Values (Columns)

☒ Deselect All

- ☒ MEASUREMENT_CONCEPT_ID
- ☒ STANDARD_CONCEPT_NAME
- ☒ STANDARD_CONCEPT_CODE
- ☒ STANDARD_VOCABULARY
- ☒ MEASUREMENT_DATETIME
- ☒ MEASUREMENT_TYPE_CONCEPT_ID
- ☒ MEASUREMENT_TYPE_CONCEPT_NAME
- ☐ OPERATOR_CONCEPT_ID

4 Preview Data Set

A visualization of your data table based on concept sets and values you selected above. Once complete, export for analysis

[View Preview Table](#)

MEASUREMENT														
STANDARD...	STANDARD...	STANDARD...	MEASURE...	MEASURE...	MEASURE...	VALUE_AS...	VALUE_AS...	VALUE_AS...	UNIT_CON...	UNIT_CON...	MEASURE...	MEASURE...	SOURCE_C...	SOURCE_C...
Systolic blood pressure	271649006	SNOMED	1970/01/16 02:13:14	44818701	From physical examination	76.0	8160311112		8876	millimeter mercury column	DIASTOLIC	4152194	Systolic blood pressure	271649006

[SAVE AND ANALYZE](#)

Save Dataset

Test Dataset

☒ Export to notebook

[HIDE PREVIEW](#)

[Python](#) [R](#)

```
import pandas
dataset_person_sql = """SELECT
person.BIRTH_DATETIME as DATE_OF_BIRTH,
person.GENDER_CONCEPT_ID, person.PERSON_ID,
person.RACE_CONCEPT_ID,
person.ETHNICITY_CONCEPT_ID,
```

(Create a new notebook)

Notebook Name

Test Notebook

Programming Language:

☒ Python

☐ R

[CANCEL](#) [SAVE AND ANALYZE](#)

Synthetic data used for training purposes

All of Us
RESEARCH PROGRAM

Analyze Data in Notebooks

A notebook is a computational environment where you can analyze data with basic programming knowledge in R or Python. Several template notebooks and resources are available within your workspace that will guide you how to export your datasets into the notebook and can assist with basic analyses.

For example, you can launch a notebook to import your “diabetes cases” cohort and then select your “biometrics” concept set, to get biometrics data for the participants in your cohort. You can then analyze the data to study correlation between hypertension and diabetes.

The screenshot shows a Jupyter Notebook interface with the title "All Participants Demographics" and a status bar indicating "Last Checkpoint: 20 hours ago (autosaved)". The notebook is running Python 3. The code in the cell is as follows:

```
In [5]: import pandas as pd

all_participants_demographics_person_sql = """SELECT person.BIRTH_DATETIME as DATE_OF_BIRTH, person.GENDER_CONCEPT_ID, person.ETHNICITY as ETHNICITY
FROM person
WHERE 1=1
"""

all_participants_demographics_person_query_config = {
    'query': {
        'parameterMode': 'NAMED',
        'queryParameters': [
            'DATE_OF_BIRTH', 'GENDER_CONCEPT_ID', 'ETHNICITY'
        ]
    }
}

all_participants_demographics_person_df = pd.read_gbq(all_participants_demographics_person_sql, dialect="standard", configuration=all_participants_demographics_person_query_config)
all_participants_demographics_person_df.head(5)
```

The output of the code is a table with 5 rows and 8 columns:

	DATE_OF_BIRTH	GENDER_CONCEPT_ID	PERSON_ID	RACE_CONCEPT_ID	ETHNICITY_CONCEPT_ID	RACE	GENDER	ETHNICITY
0	1989-01-17 00:00:00+00:00	8532	49110	8657	38003563	American Indian or Alaska Native	FEMALE	Hispanic or Latino
1	1992-12-30 00:00:00+00:00	8532	95292	8657	38003563	American Indian or Alaska Native	FEMALE	Hispanic or Latino
2	1989-01-27 00:00:00+00:00	8532	829103	8657	38003563	American Indian or Alaska Native	FEMALE	Hispanic or Latino
3	1980-04-08 00:00:00+00:00	8532	383034	8657	38003563	American Indian or Alaska Native	FEMALE	Hispanic or Latino
4	1962-06-17 00:00:00+00:00	8532	814531	8657	38003563	American Indian or Alaska Native	FEMALE	Hispanic or Latino

The next cell in the notebook is:

```
In [6]: all_participants_demographics_person_df.ETHNICITY.value_counts().to_frame()
```

The output of this cell is a table with 1 column and 1 row:

ETHNICITY