



The BROWSE & SUBSET tab on the ClinEpiDB platform



The **Browse** and **subset** tab enables you to

- Browse through a list of all the variables collected in the study
- View the univariate distributions of each of the variables
- Subset the data to choose observations of interest

It also functions as an interactive codebook.

View study details [Browse and subset](#) Visualize Notes

A **Featured variables**

- Community: Country ☆
- Household repeated measure: Household study timepoint ☆
- Participant: Case or control participant ☆
- Participant: Age group ☆
- Participant repeated measure: Study timepoint ☆
- Participant repeated measure: Age ☆

C expand all | collapse all

D Find a variable **E**

B **Variable tree**

- Community
 - Geographic location
 - Country ☆
 - Household
 - Household repeated measure
 - Participant
 - Participant repeated measure
 - Sample

Country
Original variable name: SITE [?](#)

Check items below to apply this filter

22,567 (100%) of 22,567 Communities have data for this variable

<input type="checkbox"/>	Country	Subset of Communities ?	All Communities ?	Distribution ?	% ?
		22,567 (100%)	22,567 (100%)		
<input type="checkbox"/>	Bangladesh	3,859 (17%)	3,859 (17%)	<div></div>	(100%)
<input type="checkbox"/>	India	3,582 (16%)	3,582 (16%)	<div></div>	(100%)
<input type="checkbox"/>	Kenya	3,359 (15%)	3,359 (15%)	<div></div>	(100%)
<input type="checkbox"/>	Mali	4,097 (18%)	4,097 (18%)	<div></div>	(100%)
<input type="checkbox"/>	Mozambique	1,976 (9%)	1,976 (9%)	<div></div>	(100%)
<input type="checkbox"/>	Pakistan	3,096 (14%)	3,096 (14%)	<div></div>	(100%)
<input type="checkbox"/>	The Gambia	2,598 (12%)	2,598 (12%)	<div></div>	(100%)

A. Key variables that would be useful in analyses are highlighted in a **Featured variables** box at the top of the page.

B. All variables are organized into categories (matching the dataset diagram) are listed in the **Variable tree**.

C. Click on **expand all** or **collapse all** to expand or collapse the variable tree and scroll through all variables.

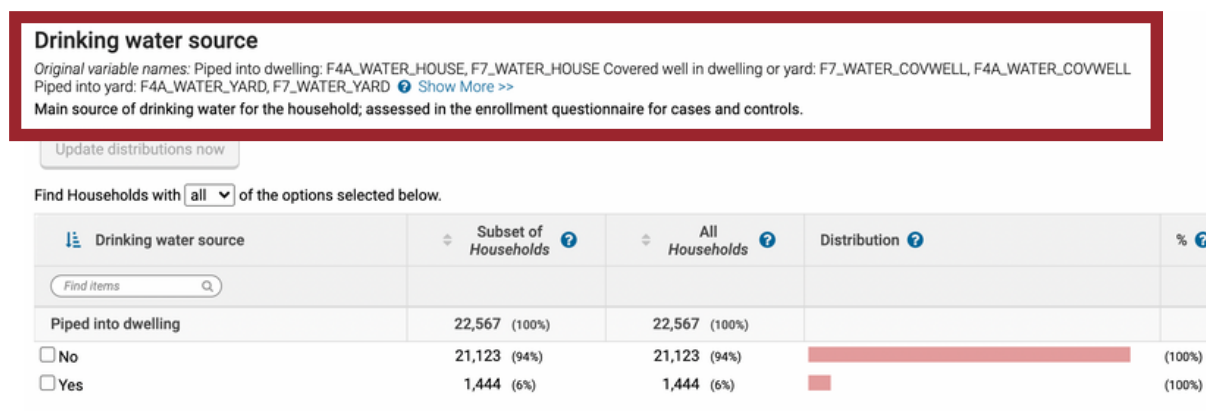
D. The **Find a variable** box allows you to search for variables.

E. Variables can be starred to make them more accessible. The toggle enables you to restrict the list to starred variables.

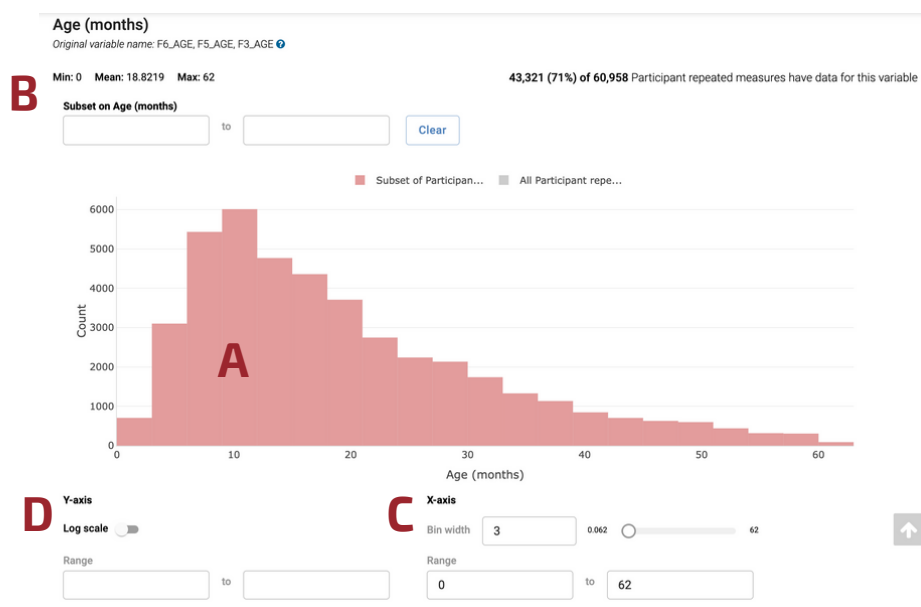
For each variable, we provide,

- **Variable labels** (Drinking water source in the example below) that are harmonized across datasets, highlighting the ontological framework underlying our data integration.
- **Original variable name** from the dataset provided by the study team
- **Definition** of the variable in many cases

All this information is grouped at the top, shown here in a red box.



Clicking on any variable label displays the distribution of the variable. In the example below, **Age (months)** is a numeric variable and its distribution is displayed with a histogram (**A**). The min, max and mean (**B**) of the variable are displayed under variable label. You can customize the appearance of the histogram by changing the bin width (**C**) or changing to log scale (**D**).



To choose a subset of numeric data, simply type the desired range within the subset box, shown below in a red box.



A subsetting variable will appear as an icon indicated by the red circle on the right, and hovering over the icon will display the chosen subset.

Browse and subset Visualize Notes

Age

Participant repeated measure: Age from 6 to 12 Original variable name: F6_AGE, F5_AGE, F3_AGE

Min: 0 Mean: 18.8219 Max: 62

Subset on Age (months)

6 to 12 Clear

The distribution of categorical variables is displayed as a frequency table (A) and a bar graph (B). In this example, **Age group** is a categorical variable with three levels.

Age group

Original variable name: agegroup

Check items below to apply this filter

22,567 (100%) of 22,567 Participants have data for this variable

	Subset of Participants	All Participants	Distribution	%
<input type="checkbox"/> Age group	22,567 (100%)	22,567 (100%)		
<input type="checkbox"/> 0-11 months	8,906 (39%)	8,906 (39%)		(100%)
<input type="checkbox"/> 12-23 months	7,586 (34%)	7,586 (34%)		(100%)
<input type="checkbox"/> 24-59 months	6,075 (27%)	6,075 (27%)		(100%)

To choose a subset of data from a categorical variable, simply tick the box or boxes of categories you want to select (red circle).

Age group

Original variable name: agegroup

Keep checked values at top

22,567 (100%) of 22,567 Participants have data for this variable

	Subset of Participants	All Participants	Distribution	%
<input checked="" type="checkbox"/> Age group	22,567 (100%)	22,567 (100%)		
<input checked="" type="checkbox"/> 0-11 months	8,906 (39%)	8,906 (39%)		(100%)
<input type="checkbox"/> 12-23 months	7,586 (34%)	7,586 (34%)		(100%)
<input type="checkbox"/> 24-59 months	6,075 (27%)	6,075 (27%)		(100%)