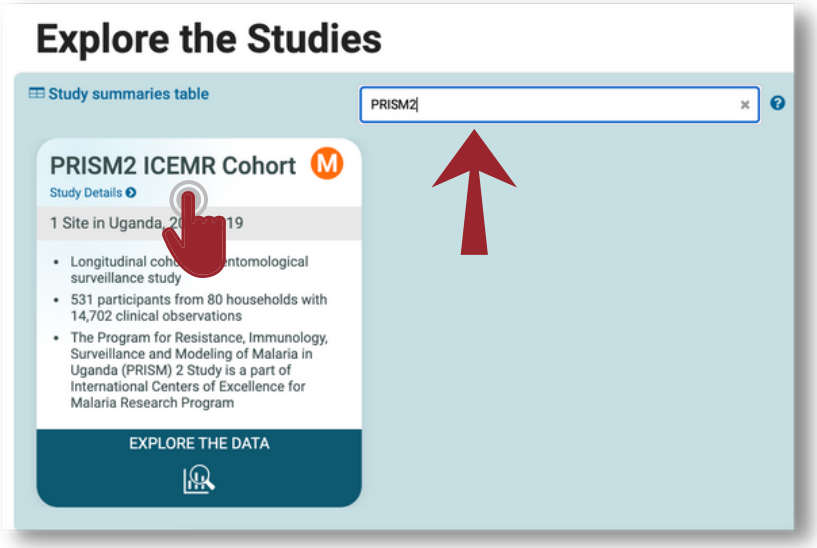
Navigating the ClinEpiDB platform
An exercise

In this exercise, we will use the **PRISM2 ICEMR Cohort study** to navigate through the different features of the ClinEpiDB platform. Type your responses into the grey boxes. Scroll to the end of the exercise for answers.

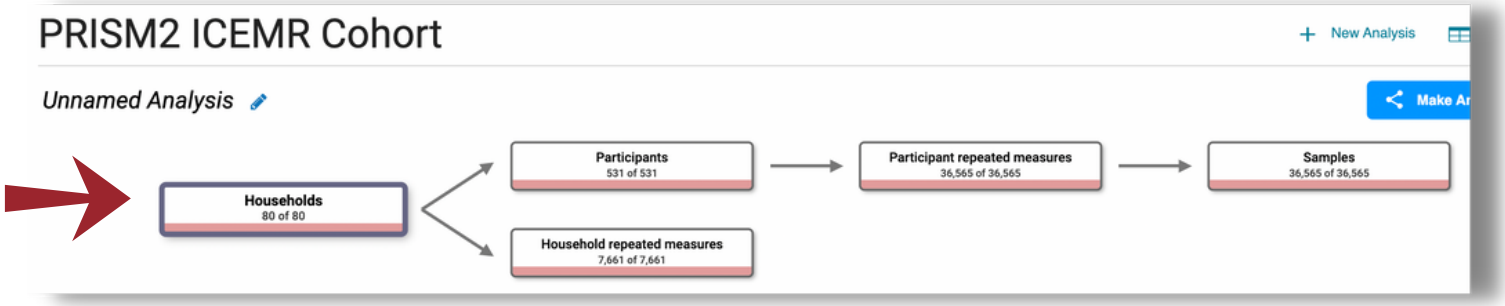
Start at the **home page**.

In the Find Studies box (red arrow), search for **PRISM2 ICEMR Cohort**.

Locate the **study card** and click on the title of the study to start exploring this dataset.



1. Look at the **Dataset diagram** (red arrow) at the top of the page and examine the shape of the data.



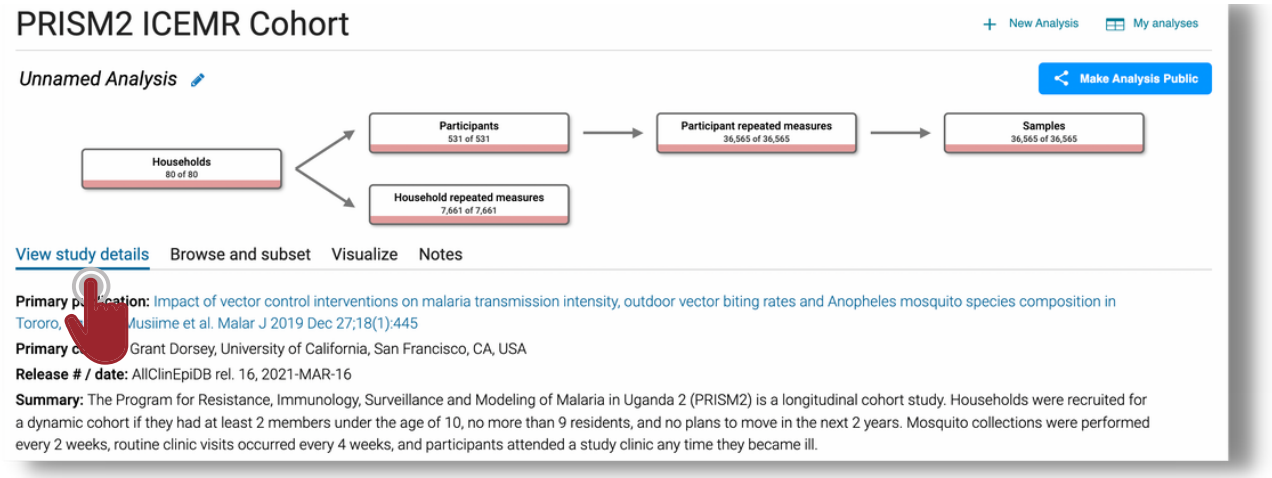
a. There are _____ participants in this study.

b. There are _____ participant repeated measures (observations) in this study.

c. What does it mean to have >1 repeated measure per participant?

_____

2. Click the **View study details** tab, scroll through the page, and answer the following:

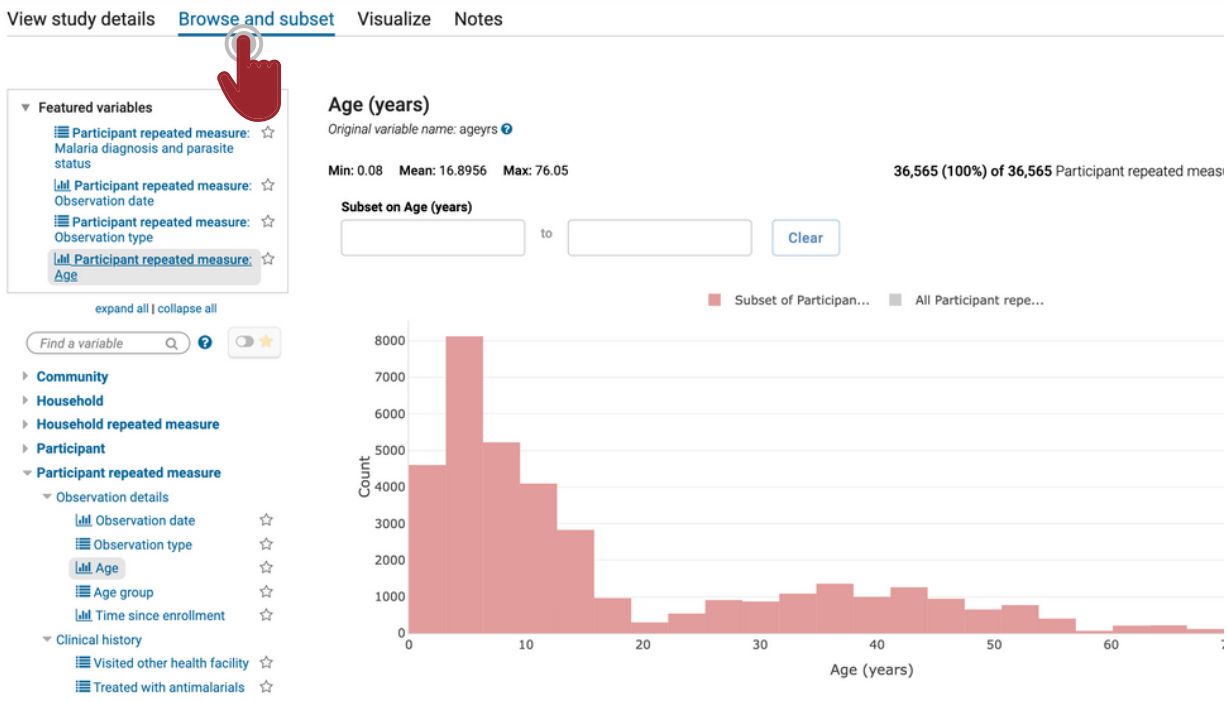


a. What is the study design?

b. What was an objective of this study?

c. Where was this study conducted?

3. Click the **Browse and subset** tab and scroll through the variable tree on the left.



a. Name one of the featured variables:

b. What proportion of participants were in the <5 years age group?

Hint: You can search for variables in the "find a variable" box in the left sidebar

c. What was the mean age of the participants?

4. Click the **Visualize** tab -> New visualization, and choose the **bar plot** tool

For the main axis variable, choose **age group**. For overlay, choose **sex.**

Under the plot, choose **proportion** instead of count.

a. What proportion of males were in the 5-15 age group?

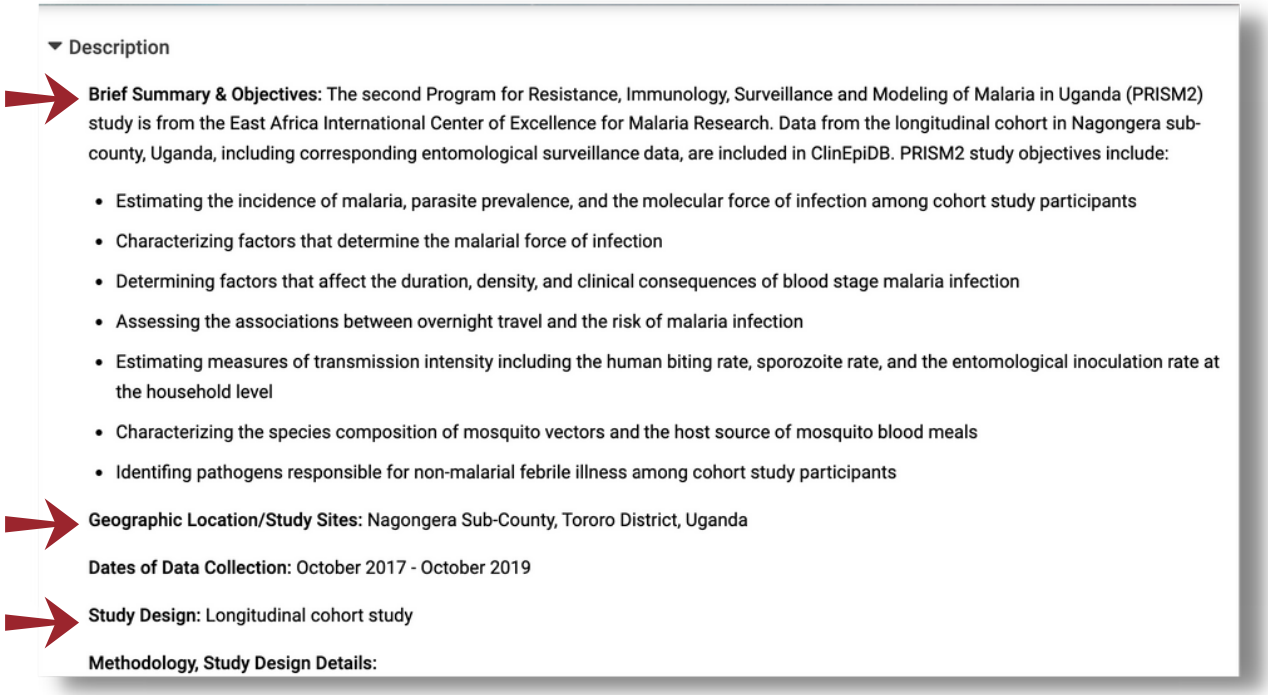
Turn to the next page for detailed answers to this exercise!

ANSWERS

1. **Dataset diagram**
    a. There are **531** participants in this study.
    b. There are **36,565** participant repeated measures (observations) in this study.
    c. What does it mean to have >1 repeated measure per participant?
    It means that participants were observed more than once.

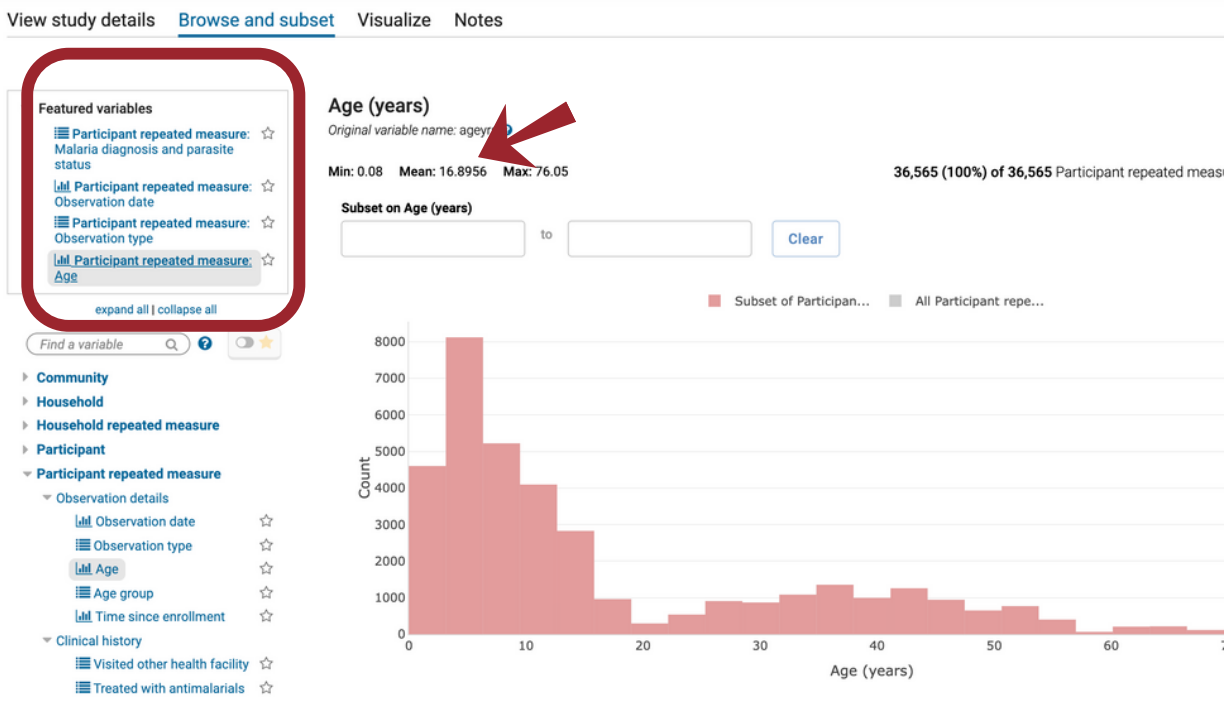
The dataset diagram at the top of the page is helpful for several reasons. It displays the various types of data collected in the study, such as data on communities, households, participants and samples, and their sample sizes. It indicates whether variables were collected once or at multiple timepoints over the study. Variables collected more than once are placed under "repeated measures". This dataset contains 36,565 repeated measures for 531 participants, which indicates that each participant was observed repeatedly over the course of the study, with an average of about 70 observations per participant.

2. **View study details** tab
    a. What is the study design? **Longitudinal cohort study**
    b. What was an objective of this study? **Estimating the incidence of malaria**
    c. Where was this study conducted? **Tororo district, Uganda**



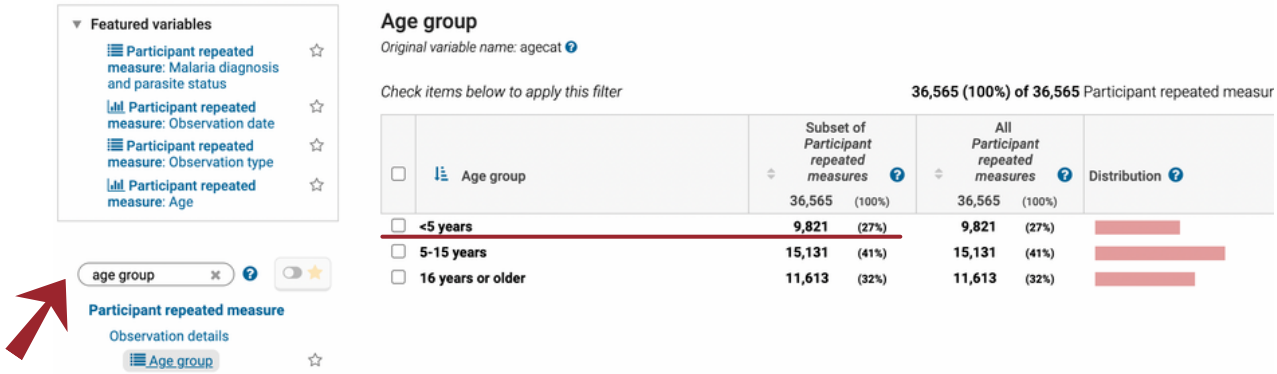
The study details page provides a wealth of information about the dataset, including links to publications, a summary of the objectives, study design and methodology, links to study documentation, and a listing of the study team.

3. **Browse and subset** tab
    a. Name one of the featured variables: **Malaria diagnosis and parasite status**
    b. What proportion of participants were in the <5 years age group? **27%**
    c. What was the mean age of the participants? **16.9 years**



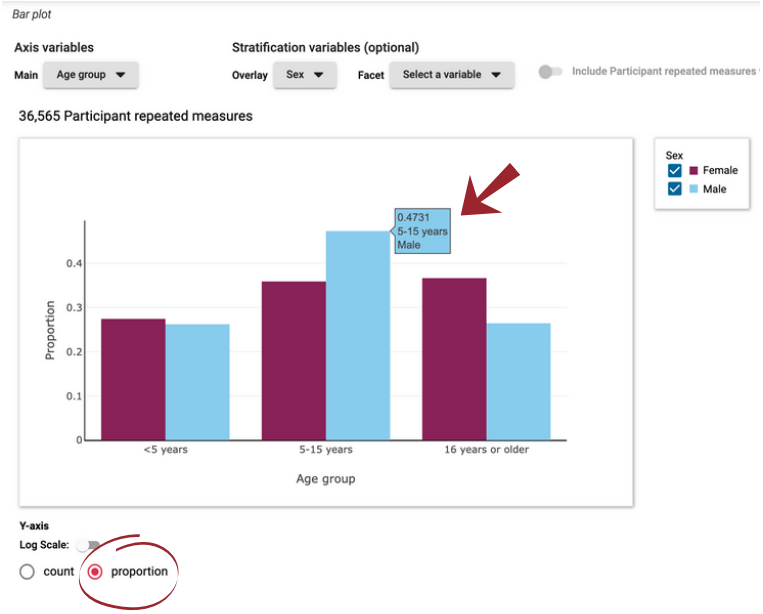
The browse and subset tab provides a list of all the variables collected in the study, organized into categories in a "variable tree" that can be expanded and collapsed. We also provide a definition for most variables, and the name of the variable in the original dataset, making this tab an interactive codebook. Key variables that would be useful in analyses are highlighted in a "featured variables" box at the top of the page. Variables can be starred to make them more accessible.

Clicking on a variable label displays the distribution of the variable. Age is a continuous variable and its distribution is displayed with a histogram. The mean age of the participants is 16.8 years, but notice that the histogram shows that age has a bimodal distribution (two peaks) representing children and adults.



You can look for the variable "age group" in the "Find a variable" search box. "Age group" is a categorical variable and its distribution is displayed with a frequency table and bar graph. We can see that 27% of the participants were in the <5 years age group.

4. **Visualize** tab
    a. What proportion of males were in the 5-15 age group? **0.473 or 47.3%**



The visualize tab provides a menu of commonly used graphs and tables to explore associations between two or more variables. In this exercise, we are graphing a single categorical variable (age group), so we chose the bar plot. By adding the overlay variable "sex", we stratified (separated) the age group by sex. We can choose to display either count or proportion. Of the males enrolled in the study (the figure legend shows that they are indicated in blue), 47.3%, or nearly half, were in the 5-15 years (school-age) group.

Thank you for completing this exercise on navigating the ClinEpiDB platform! Please contact **help@clinepidb.org** with feedback or questions.