

# Bonus: Exploring & Analyzing Enteric Epidemiology Datasets I

## Exercise 1: Getting Started - Creating a Personal EuPathDB Account and Changing your Password (10 minutes)

In this first exercise you will learn how to create a ClinEpiDB account, log in, and change your password. This exercise will be the first time you will interact directly with the site, and will give you an introductory overview of the front page.

*Register a personal account, log in, and change your password*

1. Navigate to [www.clinepidb.org](http://www.clinepidb.org) in your web browser. Take a minute to mouse-over different aspects of the front page. Notice that across the top of the homepage there is a navigation bar with different buttons. What does the mouse-over for “Search a Study” say? Don’t click on any of these tabs yet, we will mention them in more detail in the exercises to come.

Welcome To ClinEpiDB  
Advancing global public health by facilitating the exploration and analysis of epidemiological studies

### Explore the Studies

#### GEMS1 Case Control E

- Study Details E  
7 S. Asian and African Sites, 2007-2011
- The Global Enteric Multicenter Study (GEMS) investigated the causes, incidence and impact of moderate-to-severe diarrhea among children from the Gambia, Mal, Kenya, Nigeria, Uganda, Pakistan, India and Bangladesh.
  - Case-Control study with a 60-day follow-up visit
  - 16S sequence data for ~1000 stool samples available at MicrobiomeDB.org.

Download Data ▲

SEARCH THE DATA



#### GEMS1A Case Control E

Study Details E  
7 S. Asian and African Sites, 2011-2013

- The Global Enteric Multicenter Study (GEMS) 1A investigated the cause, the incidence and the impact of less-severe diarrhea (LSD).
- Case-Control study with a 60-day follow-up
- 14,242 Participants from The Gambia, Mal, Kenya, Mozambique, Pakistan, India and Bangladesh.

Download Data ▲

SEARCH THE DATA



#### India ICEMR Longitudinal Cohort Study M

Study Details M

- 2 sites in India from 2013-2015
- The Center for the Study of Complex Malaria in India (CSCMI) is part of the International Centers of Excellence for Malaria Research (ICEMR) Program
  - Longitudinal cohort study design
  - 397 participants from 110 households with 1,249 observations

Download Data ▲

SEARCH THE DATA



#### MAL-ED Study E

Study Details E

8 Sites, 2009-2014

- Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development Study
- Longitudinal birth cohort of the first 2 years of life
- Over 200 Participants from each of 8 sites (In Bangladesh, Brazil, India, Nepal, Pakistan, Peru, South Africa, Tanzania)
- 397 participants from 110 households with 1,249 observations

Download Data ▲

SEARCH THE DATA



#### PRISM E

Study Details E

3 Sites In U

- The Prog Surveillance (Uganda (International) Longitudinal birth cohort of the first 2 years of life)
- Longitudinal surveillance
- 1,421 Par with over

### News

#### ClinEpiDB 4 Released E

TUE MAR 07 2018  
We are pleased to announce the release of ClinEpiDB 4! ClinEpiDB in India March 13 - 17th, the ClinEpiDB outreach team will be attending a conference and conducting ...

#### ClinEpiDB 5 Released E

SAT DEC 15 2018  
We are pleased to announce the release of ClinEpiDB 5! New Features in this Release! The analysis apps in the Analyze Results tab of the results table have now been update... read more

#### ClinEpiDB at ASTMH E

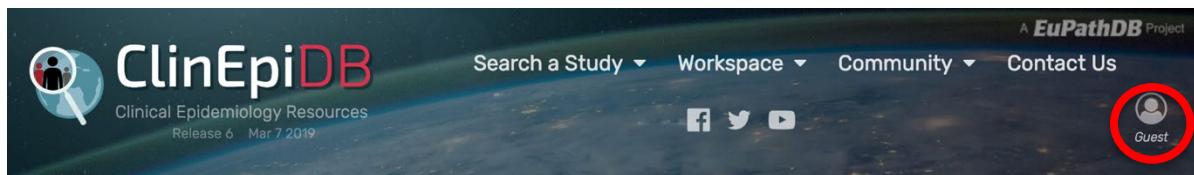
See all news

Tweets by @ClinEpiDB E

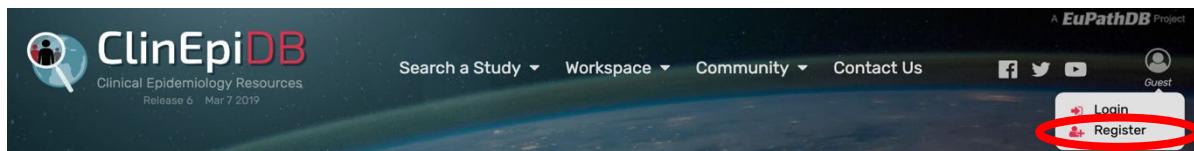


2. Each ClinEpiDB user can register for a personal EuPathDB account, protected by a Username and Password, that allows them to save and share their searches. You only need one account for ALL EuPathDB accounts. This means that if you would like to use the other EuPathDB resources such as ClinEpiDB or MicrobiomeDB, you can enter the same username and password to access saved information across sites.

At the top right-hand corner, there is an icon labeled “Guest”. Hover over this icon to access links to ‘Login’ and ‘Register’.



To register for Personal Account, click on the 'Register' link.



Fill out the registration form then click the 'Register' button. After clicking 'Register', a green text box should appear, indicating successful registration.

## Registration

Review our [EuPathDB Websites Privacy Policy](#).

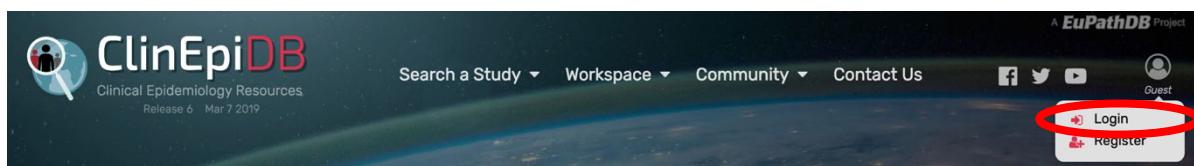
\* = required

**Identification**

\* Email: Your email will be your login username  
\* Retype Email: Please re-type the same email as above  
\* First Name:  
Middle Name:  
\* Last Name:  
\* Organization:

Check your email to retrieve your temporary password. If you are accessing your email through a web browser, open a new window or tab to access your email account. Once you receive the email message from EuPathDB, you can either write down or copy the password so you can login for the first time.

3. Return to the ClinEpiDB browser and click on the 'Login' link.



A box should pop-up. Enter the email address that you used to register and the temporary password that was emailed to you here.

## ClinEpiDB

Clinical Epidemiology Resources

Please log in

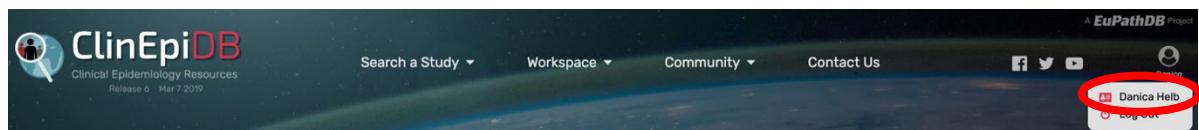
Email:   
Password:

[Forgot Password?](#) [Register/Subsribe](#)

After you complete the information and click “Login,” you will be logged into the site. You can confirm that you are logged in if your name is now listed instead of “Guest.”



You can change your password or email address by navigating to the “My Account” page. Hover over the icon to access links to “My Account” or “Log Out”. Click on your name to access your “My Account” page.



Change your password to something personalized that you will remember. Click the ‘Change your password’ link to get started.

**My Account**

Review our [EuPathDB Websites Privacy Policy](#).

\* = required

**Identification**

\* Email:

\* Retype Email:

\* First Name:

Middle Name:

\* Last Name:

\* Organization:

**Password**

[Change your password](#)

Follow the instructions on the screen and click “Submit” to change your password. If you have successfully changed your password, green text will appear confirming the password change. When you are finished, you can either click on the “ClinEpiDB” logo to return to the ClinEpiDB homepage, or you can click on the blue “click here” to return to your previous page.

**ClinEpiDB**  
Clinical Epidemiology Resources

**Change Password**

You have successfully changed your password.  
To return to your previous page, [click here](#).

Username:

Current Password:

New Password:

Re-enter New Password:

Decorative icons for various life sciences are at the bottom.

## Exercise 2: Reviewing MAL-ED study details on the “Data Set page” (10 minutes)

Before any analysis can be performed, it is imperative that you familiarize yourself with the study. You will need to know the study design, how the data were collected and structured (including timepoints for collection), and variable definitions. You also must understand the data set structure, relationships between data elements, and cardinality. Your explorations will also be aided by knowledge of how previous analyses were conducted.

This exercise will introduce you to the “Data Set page”, where you can find study procedures, methodologies, Case Report Forms (the forms used for data collection), data dictionaries, key study personnel, associated publications, and other information associated with each study. Using the resources provided on the Data Set page for the study you are interested in will provide information about the context and characteristics of the data in each study. It is always a good idea to explore the information of the Data Set page before you start exploring the data to make sure you understand the study design.

### *Reviewing study details on the Data Set page*

1. From the ClinEpiDB.org homepage, click on “Study Details” on the MAL-ED card to open the Data Set page. The Data Set page is where you will find more information about the study description, design, associated publications, investigators, and contact information for the MAL-ED study.

The image shows four study cards from the ClinEpiDB.org homepage. Each card includes a study name, a small icon, a 'Study Details' link, a brief description, a 'Download Data' button, a 'SEARCH THE DATA' button, and icons for user, download, and home.

- GEMS1 Case Control** (E)  
Study Details  
7 S. Asian and African Sites, 2007-2011
  - The Global Enteric Multicenter Study (GEMS) investigated the causes, incidence and impact of moderate-to-severe diarrhea in 22,567 children from the Gambia, Mali, Kenya, Mozambique, Pakistan, India and Bangladesh.
  - Case-Control study with a 60-day follow-up visit
  - 16S sequence data for ~1000 stool samples available at MicrobiomeDB.org.[Download Data](#) [SEARCH THE DATA](#)
- GEMS1A Case Control** (E)  
Study Details  
7 S. Asian and African Sites, 2011-2013
  - The Global Enteric Multicenter Study (GEMS) 1A investigated the cause, the incidence and the impact of less-severe diarrhea (LSD).
  - Case-Control study with a 60-day follow-up visit
  - 14,242 Participants from The Gambia, Mali, Kenya, Mozambique, Pakistan, India and Bangladesh[Download Data](#) [SEARCH THE DATA](#)
- India ICMR Longitudinal Cohort Study** (M)  
Study Details  
2 sites in India from 2013-2015
  - The Center for the Study of Complex Malaria in India (CSCMI) is part of the International Centers of Excellence for Malaria Research (ICEMR) Program
  - Longitudinal cohort study design
  - 397 participants from 110 households with 1,249 observations[Download Data](#) [SEARCH THE DATA](#)
- MAL-ED Study** (E)  
Study Details  
8 Sites, 2007-2014
  - Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development Study
  - Longitudinal birth cohort of the first 2 years of life
  - Over 200 Participants from each of 8 sites (in Bangladesh, Brazil, India, Nepal, Pakistan, Peru, South Africa, Tanzania)[Download Data](#) [SEARCH THE DATA](#)

- You can also access the MAL-ED Data Set page by hovering over “Search a Study” and clicking on “MAL-ED Study.”

The image shows the ClinEpiDB.org homepage. At the top, there is a navigation bar with links for 'PRISM Cohort Study', 'GEMS1 Case Control', 'GEMS1A Case Control', 'India ICMR Longitudinal Cohort Study', 'MAL-ED Study', and 'GEMS1 HUAS/HUAS Lite'. Below the navigation bar, there is a search bar labeled 'Search a Study' with a dropdown arrow. A red circle highlights the 'Search a Study' button. To the right of the search bar, there are links for 'Workspace', 'Community', and 'Contact Us', along with social media icons for Facebook, Twitter, and YouTube. The background features a dark image of a city skyline at night with the text 'Welcome To ClinEpiDB' and 'Enabling the exploration and analysis of epidemiological studies'.

2. Scroll down the MAL-ED Data Set page to see what types of study details are available.

The screenshot shows the ClinEpiDB interface for the MAL-ED Study. At the top, there are navigation links: Search a Study, Workspace, Community, Contact Us, and a user profile. Below the header, the title "Data Set: MAL-ED Study" is displayed. A summary box provides an overview of the study's purpose, design, and key findings. The "Associated Publications" section lists numerous academic papers from various countries and years, detailing the study's methodology and results. Contact information for David Spiro and the Fogarty International Center is also present.

3. Use the information on the MAL-ED Data Set page to answer the following questions:

- **Question 1: At what ages were participants enrolled in the MAL-ED study?**
- **Question 2: At what ages were blood samples collected from study participants?**
- **Question 3: Where can you download the original Case Report Forms used in the MAL-ED study?**
- **Question 4: Where can you find a link out to all MAL-ED studies on PubMed?**

## **Answers:**

### **Question 1: At what ages were participants enrolled in the MAL-ED study?**

- Answer: Participants were enrolled within the first 17 days of life.

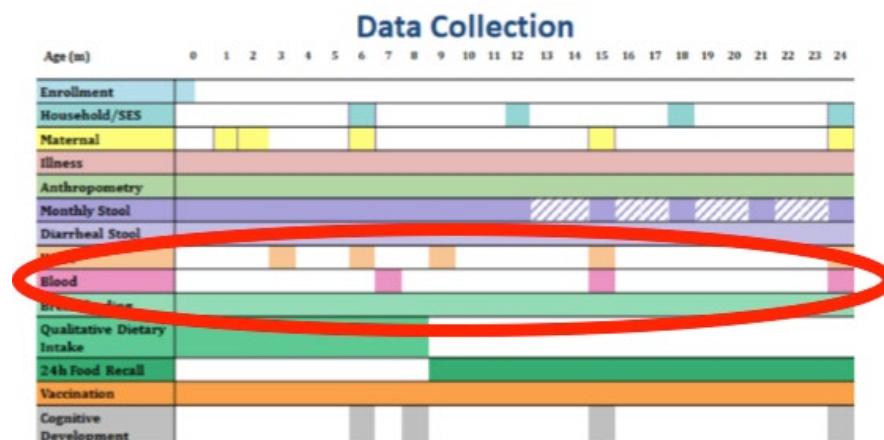
**Methodology, Study Design Details:** In contrast to cross-sectional and case-control studies, the MAL-ED study was designed as a observational study that used a prospective longitudinal design to directly address a complex system of exposures and health outcomes in children born and living in eight culturally and geographically diverse resource-constrained communities in low and middle income countries. Children were enrolled within the first 17 days of life and were visited twice weekly through to 24 months of age to collect data as represented in the figure below. Each site recruited more than 200 children over a 24-month time period – targeting a smooth and constant monthly enrolment across this time frame – to assess seasonal effects of exposures. Details of the study design and descriptions of data and sample collection have been published (*Clinical Infectious Diseases*, vol. 59, suppl. #4, November, 2014).

Subjects were enrolled after signed informed consent was obtained according to appropriate procedures agreed to for this project. All sites used a common standardized protocol that was developed and agreed to by study investigators, and was reviewed and approved by national, local and organizational review boards, as appropriate and required, for each collaborating institution.

The aim of the MAL-ED study is to improve scientific understanding of the complex interrelationships among enteropathogen infection, dietary intake, nutritional status, gut physiology, growth, immune function and vaccine response, and cognitive development. The timing of data and sample collection and the assessments determined are shown here.

### **Question 2: At what ages were blood samples collected from study participants?**

- Answer: Blood was collected from study participants at 7, 15, and 24 months of age.



### **Question 3: Where can you download the original Case Report Forms used in the MAL-ED study?**

- Answer: Scroll down to the “Case Report Forms and Data Dictionaries” section. There, you can find a link to download the “Case report form codebook”. The original case report forms are contained in this download file.

#### **Case Report Forms and Data Dictionaries:**

- [Case report form codebook](#) – The codebook for each case report form is included as a separate sheet
- [Analytic dataset codebook](#) – The codebook for each analytic dataset is included as a separate sheet

### **Question 4: Where can you find a link out to all MAL-ED studies on PubMed?**

- Answer: Scroll down to “Publications from the project can be found by clicking here” and click on the “PubMed Search for MAL-ED” link. Note that there are also links to key publications in the “Associated Publications” section, but this is not a comprehensive list of all MAL-ED publications.

Publications from the project can be found by clicking here: [PubMed Search for MAL-ED](#)

## Exercise 3: Accessing and Exploring MAL-ED Data via the Filters – Guided Walk-Through on using ClinEpiDB (40 minutes)

In this exercise, we will walk through how to explore data on ClinEpiDB.org together, using the MAL-ED study as an example. This will be your first hands-on introduction on accessing data on the website. Make sure to click along on your own computer as we go through this exercise.

### Types of searches: Participant, Observation, Household, and Light Traps

1. We are now going to explore MAL-ED study data. Hover over “Search a study” to access links to the different studies loaded into ClinEpiDB.

The screenshot shows the ClinEpiDB homepage. At the top, there is a navigation bar with links for 'Workspace', 'Community', 'Contact Us', and a user profile. Below the navigation bar, there is a search bar labeled 'Search a Study' with a magnifying glass icon. A red circle highlights this search bar. To the right of the search bar, there are three study cards: 'India ICEMR Longitudinal Cohort Study' (with a 'M' icon), 'MAL-ED Study' (with a 'E' icon), and another 'India ICEMR Longitudinal Cohort Study' (with a 'M' icon). Each card has a 'Study Details' link, a 'Download Data' button, and a 'SEARCH THE DATA' button with icons for participant (person), observation (person with a house), household (house), and light trap (insect). On the right side of the page, there is a 'News' section with several news items and a 'Tweets by @ClinEpiDB' section.

Take a look at the black icons that are located by “MAL-ED Study”. Notice that, for the MAL-ED study, there are two types of searches you can execute on the data:

- i) Participant
- ii) Observation

The Participant search and the Observation search will return two different levels of data:

- By conducting a Participant-level search, you will be examining and potentially filtering the data on characteristics or observations about the participants, but you will only retrieve one row of data per participant.
  - For MAL-ED, “Participants” are any children who were enrolled in the study.
  - There may be data within this dataset that refers to the child’s mother or caregiver but mothers, fathers, or caregivers are not considered participants in the MAL-ED study. This data is only available as associated with the child.
- When conducting an Observation-level search, you will also be examining and filtering the data based on characteristics or observations about the participants, but instead you may retrieve many rows of data per participant. Each row returned represents the data collected for each participant on a given observation date.

Note that there are no icons for Household-level or Light Trap-level searches. The MAL-ED study only enrolled one participant per household and did not collect light trap (insect) data.

## *Initiate a Participant-level search*

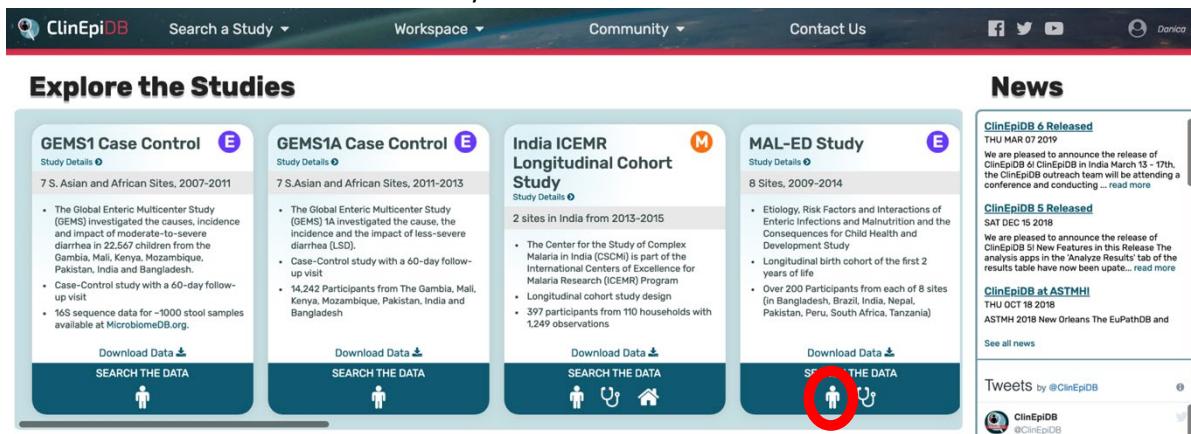
2. Let's get started and conduct a Participant-level search on the MAL-ED study so you can see what the resulting data table will look like. You can access the Participant-level search in three ways.

- Through the navigation bar at the top of the page. Hover over "Search a Study", and click on the "Participants" icon for the MAL-ED study.



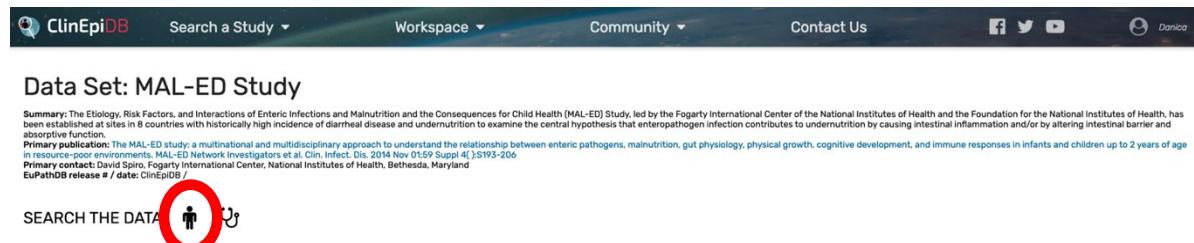
The screenshot shows the ClinEpiDB homepage. At the top, there is a navigation bar with links for 'Search a Study', 'Workspace', 'Community', and 'Contact Us'. Below the navigation bar, there is a section titled 'Explore' featuring several study cards. One of the study cards is for the 'India ICEMR Longitudinal Cohort Study' (Study Details). To the right of the study cards, there is a 'News' section with a box for 'ClinEpiDB 6 Released'. At the bottom of the page, there is a 'News' sidebar with recent news items and a 'Tweets by @ClinEpiDB' section.

- Through the MAL-ED study card. Click on the "Participants" icon under "Search the Data" on the MAL-ED study card.



The screenshot shows the 'Explore the Studies' page on ClinEpiDB. It features four study cards: 'GEMS1 Case Control', 'GEMS1A Case Control', 'India ICEMR Longitudinal Cohort Study', and 'MAL-ED Study'. The 'India ICEMR Longitudinal Cohort Study' card is currently selected. Below the study cards, there is a 'News' sidebar with recent news items and a 'Tweets by @ClinEpiDB' section.

- Through the MAL-ED Data Set page. Click on the "Participants" icon next to "Search the Data" near the top of the MAL-ED Data Set page.



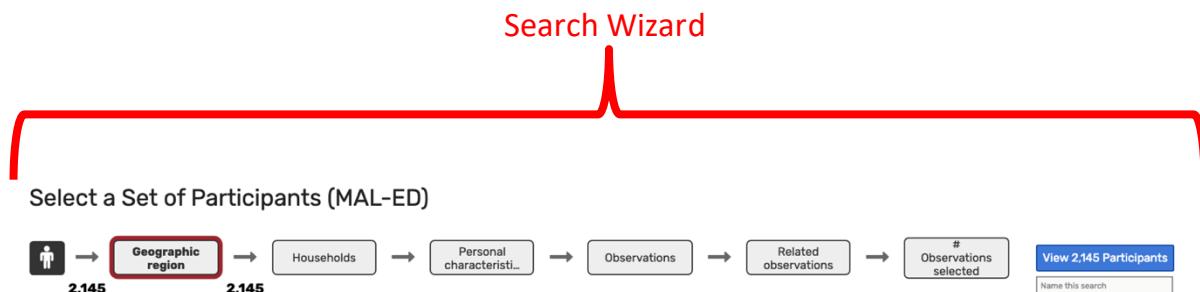
The screenshot shows the 'Data Set: MAL-ED Study' page on ClinEpiDB. At the top, there is a navigation bar with links for 'Search a Study', 'Workspace', 'Community', and 'Contact Us'. Below the navigation bar, there is a summary of the study and a primary publication. At the bottom of the page, there is a 'SEARCH THE DATA' button with a 'Participants' icon next to it. There is also a 'Tweets by @ClinEpiDB' section at the very bottom of the page.

---

### The Search Wizard

---

3. All three methods to access the Participant-level search will bring you to the same search page, which includes the “Search Wizard”.



The purpose of the search wizard is two-fold.

- i. It creates a simple way to categorize components of the data, allowing for a step-wise approach to building searches.
- ii. It allows you to explore the data to see what the raw values and distributions of characteristics are in both the full dataset or filtered data.

Spend some time reading and mouse over the different boxes in the Search Wizard. Notice that there is a number below the black Participant icon. This number represents the total number of Participants that are included in this dataset. The MAL-ED study currently has 2,145 participants (children) and their associated data loaded in this version of ClinEpiDB.

---

### Finding variables and examining data in the variables

---

4. Click on the “Geographic Region” box. What happens when you click?<sup>1</sup> Without applying a filter, can you tell which country had the greatest number of participants?<sup>2</sup>

Select a Set of Participants (MAL-ED)



<sup>1</sup> Text appears that says, “No Geographic Region filters applied yet” and a “Countries” table is displayed. The table has columns for country, remaining participants, all participants, and distributions of the counts of participants from each country.

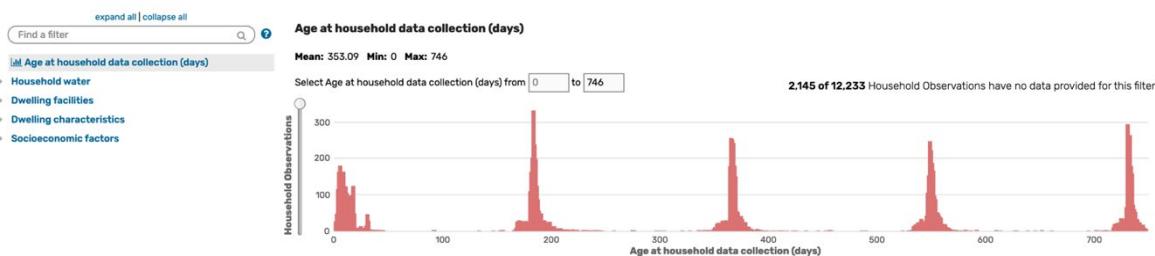
<sup>2</sup> South Africa had the greatest number of participants.

5. Next, click on the “Households” box. This will reveal a more complex filtering table that includes major categories of variables in blue text on the left and summary information for the selected variable on the right.

#### Select a Set of Participants (MAL-ED)



By default, “Age at household data collection (days)” is highlighted and an interactive histogram appears. This variable enables you to choose particular timepoints of data collection for repeated household measurements. Examine the distribution of “Age at household data collection (days)” in the histogram. About how many days apart are each of these peaks? Why do you think there are peaks occurring at regular intervals?<sup>3</sup>



6. Click “Expand all” to see the organizational hierarchy of the Household variables and reveal all the types of data in each category and sub-category.



7. Notice that categories and sub-categories have an arrow to their left, while variables do not. What happens when you click on a blue category label? What happens when you click on a blue variable label?<sup>4</sup>

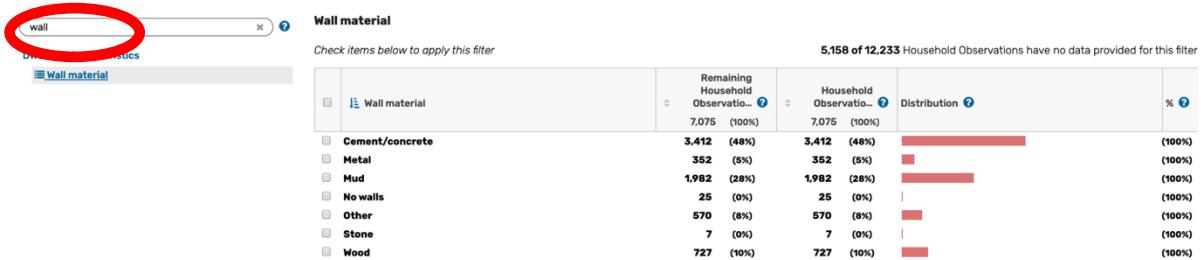
8. Spend some time clicking through the variable filters under the “Households” box in the Search Wizard and examining the distribution of participant data in each.

---

<sup>3</sup> The peaks are approximately 180 days apart. Household observations were collected at multiple time points during the MAL-ED study, and had a sampling frame of ~6 months.

<sup>4</sup> When you click on a category label, the organizational hierarchy of that category either expands or collapses. When you click on a variable label you open up a table of data that corresponds to the variable label you selected.

9. Start typing “wall” into the “Find a filter” search box to quickly navigate to the “Wall material” variable, then click on the blue “Wall material” variable label to see the data for all participants. Notice that when you mouse over a row in the “Wall material” table, it will be highlighted in a grey color to help you read the table.



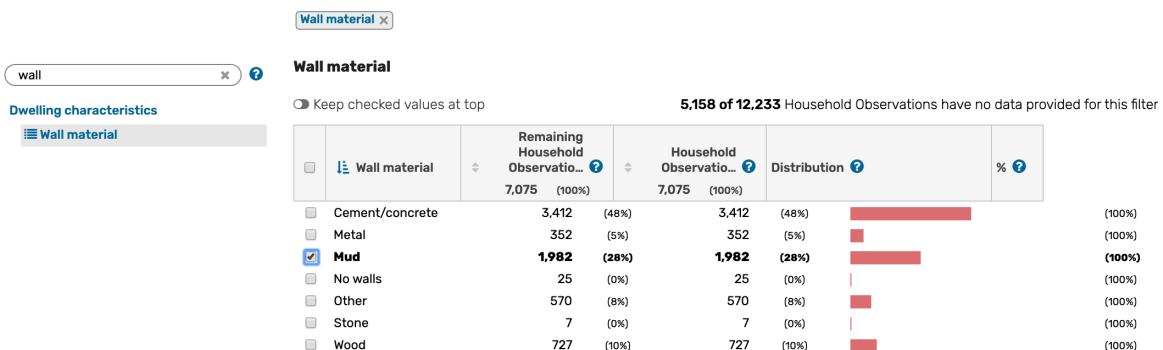
10. How many participants indicated that they lived in houses with mud walls?<sup>5</sup> Hint: 1,982 indicates the number of OBSERVATIONS where participants lived in houses with mud walls, NOT the number of PARTICIPANTS who live in houses with mud walls.

11. Can you find the information about ‘Drinking Water Source’? How many participants indicated that they had a tube well or borehole as their water source?<sup>6</sup> Hint: make sure that you have removed the “Wall material” filter applied in Step 8.

<sup>5</sup> 642 participants live in houses with mud walls. Notice that when you click to select “Mud”, the number of participants in the Search Wizard is reduced from 2,145 to 642. The wall material of the participants houses was assessed several times over the course of the longitudinal MAL-ED study, and 1,982 indicates the total number of observations where the house was assessed as having mud walls.



Your Households filters reduce 2,145 Participants to 642



<sup>6</sup> 551 participants had a tube well or borehole as their drinking water source. NOTE: if you have determined that 294 participants had a tube well or borehole as their drinking water source, your search still has the “Wall material” filter applied. To remove this filter, click on the “X” in the grey box above the table.

## Removing filters that have been applied to the data

12. Before proceeding, remove any filters that you may have applied in previous steps. This can be done in two different ways:

1. If filters have been applied to the data, a filter icon will appear in one or more of the boxes in the Search Wizard and there will be a grey box indicating what filter has been applied. To remove the filter, click on the “X” on the right-hand side of the grey box.

The screenshot shows the Data Explorer search wizard. The flow is: Participant → Geographic region → Households (with a green filter icon) → Personal characteristics → Observations → Related observations → # Observations selected → View ? Participants. The 'Households' step shows a count of 2,145 reduced to 551. A red circle highlights the green filter icon next to 'Households'.

**Your Households filters reduce 2,145 Participants to 551**

**Drinking water source** (highlighted with a red circle)

Remaining Household Observatio... ?	Household Observatio... ?	Distribution ?
7,077 (100%)	7,077 (100%)	
380 (5%)	380 (5%)	
1,396 (20%)	1,396 (20%)	
1,419 (20%)	1,419 (20%)	
113 (2%)	113 (2%)	
1,273 (18%)	1,273 (18%)	
367 (5%)	367 (5%)	
<b>1,692 (24%)</b>	<b>1,692 (24%)</b>	
437 (6%)	437 (6%)	

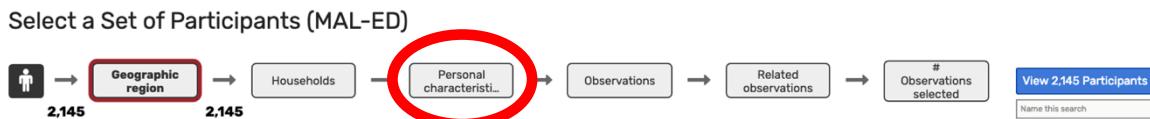
2. A green filter icon also appears next to “Select a Set of Participants”. Click on this green filter icon to see all of the filters that have been applied to the data.

Select a Set of Participants (MAL-ED)

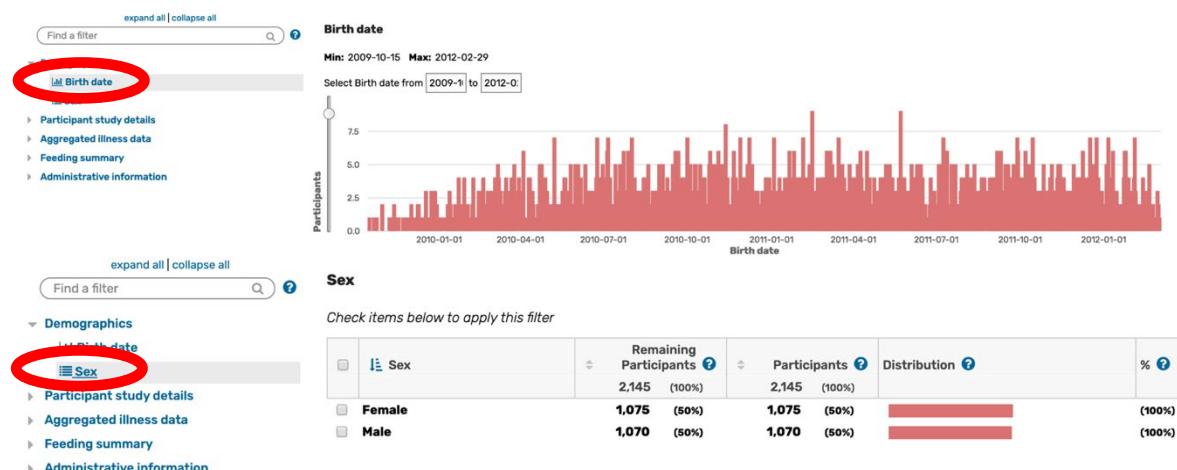
A pop-up box appears. Click on “Remove all” on the bottom left side of the pop-up box to remove all filters.



13. Next, click on the ‘Personal Characteristics’ box in the search wizard. For longitudinal studies, variables under ‘Personal Characteristics’ are generally those that are static for a given participant (such as birth date or sex) or summarize data collected over the course of the study. Here, “Birth date” is the default variable displayed. Please note that all dates in ClinEpiDB have been randomly obfuscated by +/- 7 days in a consistent manner for each participant to ensure that no identifying information is available on the website.



14. Explore the “Demographics” variables. You may notice that the figure on the right-hand side next to the filter is an interactive histogram for “Date of Birth,” while a table appears for “Sex”. Why might this be?<sup>7</sup>



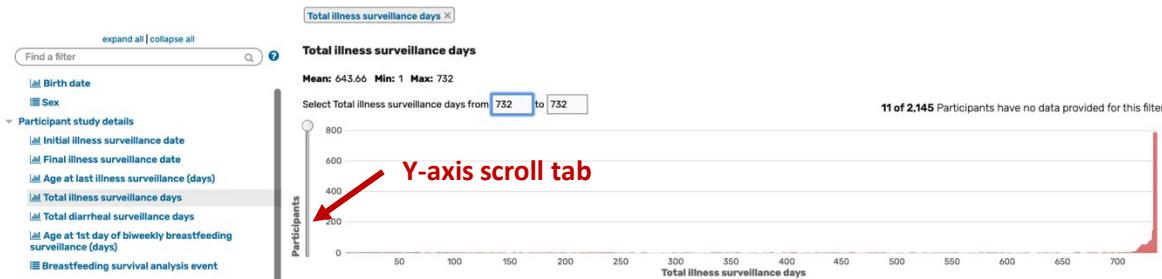
15. Navigate to the “Total illness surveillance days” variable. Remember that so far we have not filtered our participants by any criteria. Notice that the histogram on the right has an x-axis range of zero to above 700. Can you find the maximum number of days that a participant was followed up? How about the average?<sup>8</sup>

<sup>7</sup> Categorical variables are displayed in table format. Numerical variables having more than possible 10 values are displayed as a histogram rather than in table format. This allows you to see the distribution of values of your selected participants.

<sup>8</sup> These statistics are displayed above the histogram. The average was 643.66 and the maximum was 732 days of follow-up. This information is confirmed in the histogram because you can see that there are very few counts of participants towards the left end of the x-axis and increasing counts towards the right end of the x-axis.

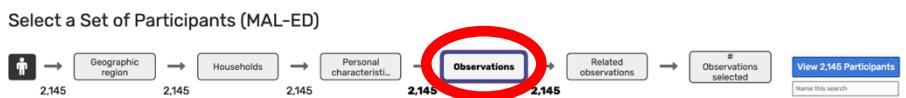


16. Can you tell how many participants had the largest number of days of illness surveillance?<sup>9</sup>  
*Hint: Use the scroll tab on the y-axis to adjust the scale of the y-axis for an approximate answer, or use the “Select Total illness surveillance days” search boxes for an exact answer.*



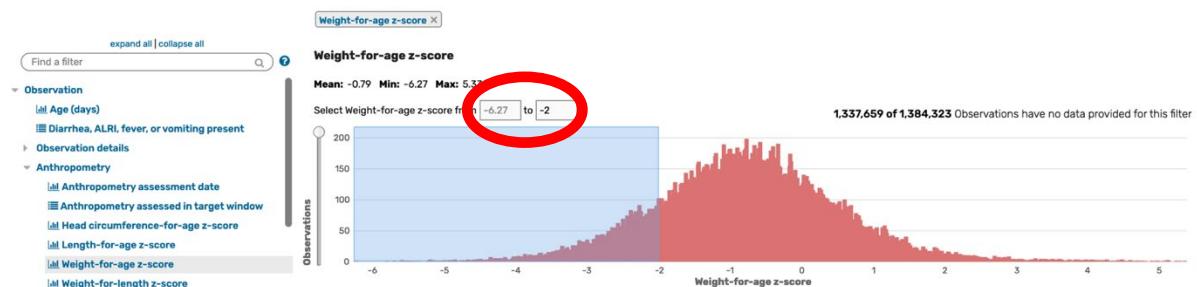
17. Click through the different filter categories found in the personal characteristics section. Notice that some filters are displayed categorically and some are displayed with distribution histograms. How many Participants were breastfed within the first 24 hours after birth?<sup>10</sup> Remember that you must remove the “Total illness surveillance days” filter you applied in the last step.

18. Click on the ‘Observations’ box in the search wizard. The default filter that is displayed shows a histogram of the age in days when an observation occurred.



19. While you are still in the ‘Observations’ section, navigate to “Weight-for-age z-score” and select children who had a weight-for-age z-score of less than -2. When you select participants who had a weight-for-age z-score measurement of less than or equal to -2, how many participants remain?<sup>11</sup> Note that you can select continuous variables displayed in a histogram in two ways:

- Type the exact number of days in the selection boxes that are directly above the histogram. Note that once you type in the desired range you must click enter on your keyboard to apply this filter. The histogram should dynamically react and display the selected day range.
- Click on the histogram around -2 days and drag your cursor left. You will notice that the histogram turns white and now only your selection is blue. Because the units are so small, you may need to adjust your selection using the boxes described above to get the selection exact.



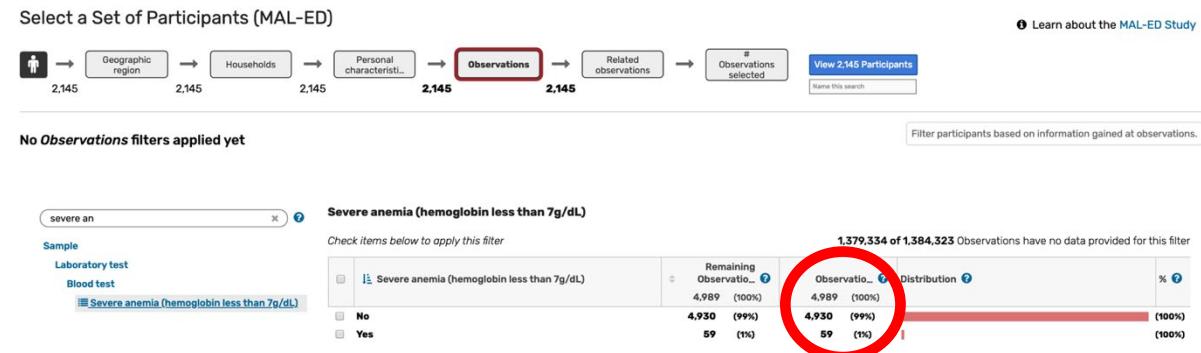
<sup>9</sup> There were 787 participants who were followed for the maximum number of days (732 days). Notice that the number of Participants has been reduced from 2145 to 787 after applying the “Total illness surveillance days” filter.

<sup>10</sup> 1995 participants were breastfed within the first 24 hours after birth.

<sup>11</sup> 847 participants had at least 1 recorded weight-for-age z-score measurement of less than or equal to -2.

20. Next, remove all active filters by clicking on the “X” in the grey boxes that appear in the “Active Filters” pop-up. Alternatively, you could click on “Remove all” at the bottom of the “Active Filters” pop-up. Notice that the number of participants goes back up to 2145.

21. While you are still in the ‘Observations’ section, navigate to “Severe anemia (hemoglobin less than 7g/dL)”. Notice that there are now 4,989 Observations available. Why are there more observations than the 2145 MAL-ED Participants?<sup>12</sup>



22. While still looking at the “Severe anemia (hemoglobin less than 7g/dL)” variable, click to check the box to the left of “Yes”. By checking the box for “Yes,” only participants that have an associated observation of severe anemia are selected for inclusion.

- Why is it that only 53 participants remain after selecting participants who had severe anemia and not 59 participants?!?<sup>13</sup>
- How many MAL-ED participants did NOT have severe anemia observed over the course of study?<sup>14</sup>

23. Uncheck the box for “Yes” again. Notice that the grey box above the filters disappears and the sentence below the search wizard again reads “No Observations filters applied yet”.

<sup>12</sup> MAL-ED is a longitudinal study, and several of the variables listed under “Observations” were measured more than one time for each participant.

<sup>13</sup> There were 59 **observations** of severe anemia in the MAL-ED study, but 6 participants had more than 1 recorded hemoglobin level less than 7g/dL. Therefore, severe anemia was recorded in only 53 participants.

<sup>14</sup> 2092 participants in the MAL-ED ClinEpiDB database did not report having severe anemia during their follow-up period and have been removed from further analysis.  $2145 - 53 = 2092$ .

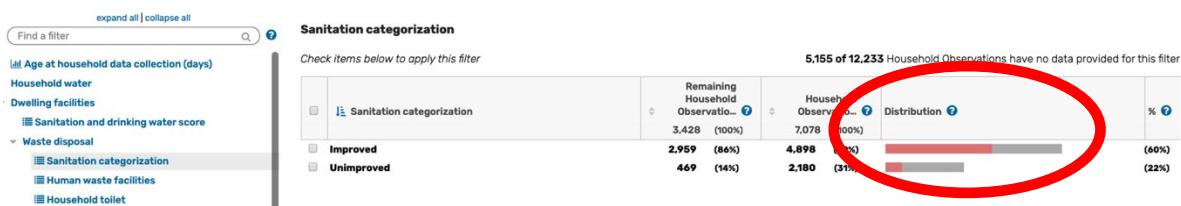
---

*Examine the impact of selecting data on other variables*

---

24. Let's use ClinEpiDB.org to determine the percentage of Household Observations with improved sanitation located in urban study sites. This requires interacting with two different variables.

- First, initiate a new participant-level search of the MAL-ED data. Next, click on the “Geographic region” box on the Search Wizard and navigate to the “Urban or rural site” variable. Select participants from “urban” sites. When you select participants from urban sites, how many participants remain?<sup>15</sup>
- Now navigate to “Sanitation categorization” under “Household observations”. Notice that part of the Distribution bar graph is now in grey and part is in red. The red portion represents data from participants who remain, while the grey represents the participants who did not meet the filter criteria you have applied. In this case, red represents participants from urban sites and grey represents rural participants.



- What percentage of household observations across both urban and rural sites had a sanitation categorization of “improved”?<sup>16</sup> What about urban sites only?<sup>17</sup>

25. Click the box to select participants with a “Sanitation categorization” of “improved.” Note that there are 824 participants from urban sites who live in houses with improved sanitation.

Note: You can always click on an earlier step in the Search Wizard and add or remove filters. However, filters are only applied in the order of the Search Wizard, so if you filter based on a variable under “Observations”, you will not see the data under “Personal characteristics” update to reflect that.

---

<sup>15</sup> 989 participants from urban sites remain.

<sup>16</sup> The numbers under the “Household observations” column represent ALL of the data. Thus, across all sites (both urban and rural), 69% of household observations had a sanitation categorization of “improved”.

<sup>17</sup> The numbers under the “Remaining Household observations” column indicate the number of participants that match the filters that you have applied. In this case, 86% of household observations conducted at urban sites had a sanitation categorization of “improved”.

## Viewing results

26. Click on the blue “View 824 Participants” box in the Search Wizard to bring up the results table of selected participants.



**824 MAL-ED 0-24 Participants** [Revise](#) [Combine with another search](#) [Save](#) [Share](#)

MAL-ED 0-24 Participant Results [Analyze Results](#)

	Participant Id	Country	Total illness surveillance days	Total diarrheal episodes	Total ALRI episodes	Age last exclusively breastfed (days)
<a href="#">e2c37aa81b8851f3</a>	Brazil	168	0	0	29	
<a href="#">d788b2acbef3cd16</a>	Bangladesh	180	2	1	139	
<a href="#">e76876f812c9f00</a>	Brazil	184	0	0	23	
<a href="#">2ff43e2f3e6f73aa</a>	Brazil	184	0	0	30	
<a href="#">b9d77ae9c214616c</a>	Nepal	191	2	1	24	
<a href="#">64cbec783484f4e6</a>	Brazil	190	0	0	124	
<a href="#">20dcc57dadce43b</a>	Brazil	206	0	0	26	
<a href="#">8ff625af2085e5c1</a>	Nepal	215	0	0	130	
<a href="#">87524fcabb935e66</a>	Brazil	208	0	0	4	
<a href="#">1856a59e97b3979b</a>	Brazil	203	0	0	38	
<a href="#">306e21007be591f9</a>	Nepal	218	0	0	57	
<a href="#">9c48a08abe067c74</a>	Bangladesh	222	4	2	2	
<a href="#">c8524ec671bb65f4</a>	Bangladesh	233	1	0	105	

27. The Results Table has one row for each of the 824 Participants you selected by performing the participant-level search above. What do you think each row would represent if you had performed an observation-level search instead of a participant-level search?<sup>18</sup>

28. You can choose to include additional variables in the results table by clicking on the “Add Columns” button on the upper right side of the table. This will bring up a popup window where you can select the variables you would like to have displayed in the table.

**824 MAL-ED 0-24 Participants** [Revise](#) [Combine with another search](#) [Save](#) [Share](#)

MAL-ED 0-24 Participant Results [Analyze Results](#)

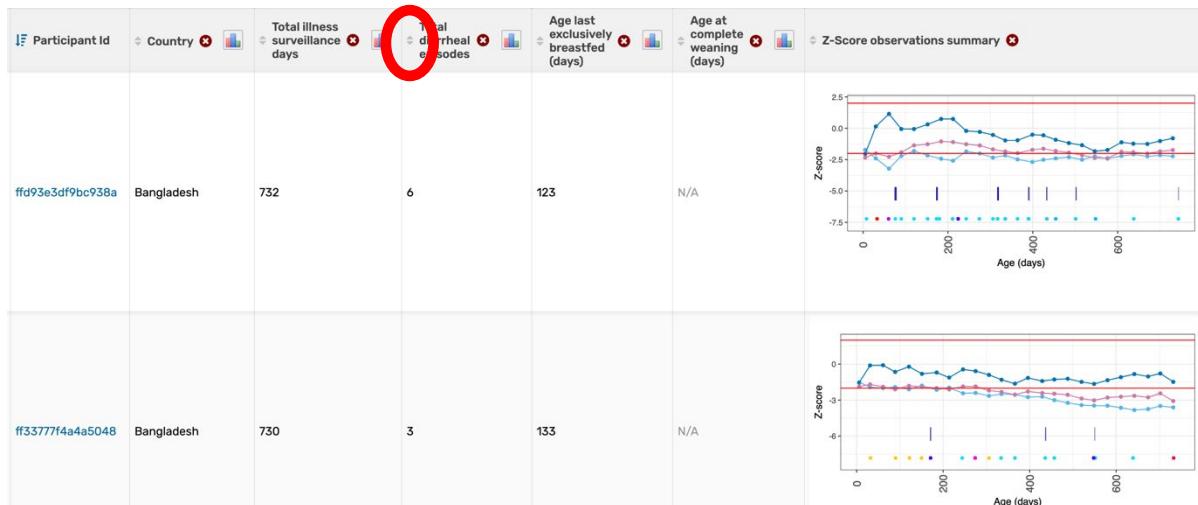
	Participant Id	Country	Total illness surveillance days	Total diarrheal episodes	Total ALRI episodes	Age last exclusively breastfed (days)
<a href="#">e2c37aa81b8851f3</a>	Brazil	168	0	0	29	
<a href="#">d788b2acbef3cd16</a>	Bangladesh	180	2	1	139	
<a href="#">e76876f812c9f00</a>	Brazil	184	0	0	23	
<a href="#">2ff43e2f3e6f73aa</a>	Brazil	184	0	0	30	
<a href="#">b9d77ae9c214616c</a>	Nepal	191	2	1	24	
<a href="#">64cbec783484f4e6</a>	Brazil	190	0	0	124	
<a href="#">20dcc57dadce43b</a>	Brazil	206	0	0	26	
<a href="#">8ff625af2085e5c1</a>	Nepal	215	0	0	130	
<a href="#">87524fcabb935e66</a>	Brazil	208	0	0	4	
<a href="#">1856a59e97b3979b</a>	Brazil	203	0	0	38	
<a href="#">306e21007be591f9</a>	Nepal	218	0	0	57	
<a href="#">9c48a08abe067c74</a>	Bangladesh	222	4	2	2	
<a href="#">c8524ec671bb65f4</a>	Bangladesh	233	1	0	105	

<sup>18</sup> If you had initiated an observation-level search on the MAL-ED data instead of a participant-level search, each row in the Results Table would represent data from a particular observation date for a participant. Each unique participant would be represented by more than one row in the Results Table. The number of rows for each participant would be equal to the number of days when data was collected from that participant.

29. In the “Select Columns” popup, click to select: (i) “Country,” (ii) “Total illness surveillance days,” (iii) “Total diarrheal episodes,” (iv) “Age last exclusively breastfed (days),” (v) “Age at complete weaning (days),” and “Z-score observations summary.” Deselect all other variables. Click the “Update Columns” button in the popup window to update the Results Table so that the variables you have selected appear as columns in the Results Table.

The screenshot shows the ClinEpiDB interface. On the left, the 'Results Table' displays data for 824 participants, with columns for Participant ID, Country, Total illness surveillance days, and Total diarrheal episodes. A red circle highlights the 'Update Columns' button in the 'Select Columns' dialog box. On the right, a chart titled 'Z-Score observations summary' plots Z-score against Age (days), showing a fluctuating line with vertical error bars and colored dots representing individual data points.

30. You can sort the table by clicking on the grey arrows to the left of the variable name in the Results Table column header. Sort the table so that the participants with the greatest number of diarrheal episodes are at the top. How many of the 824 selected participants had more than 20 diarrheal episodes over the first 2 years of study?<sup>19</sup>



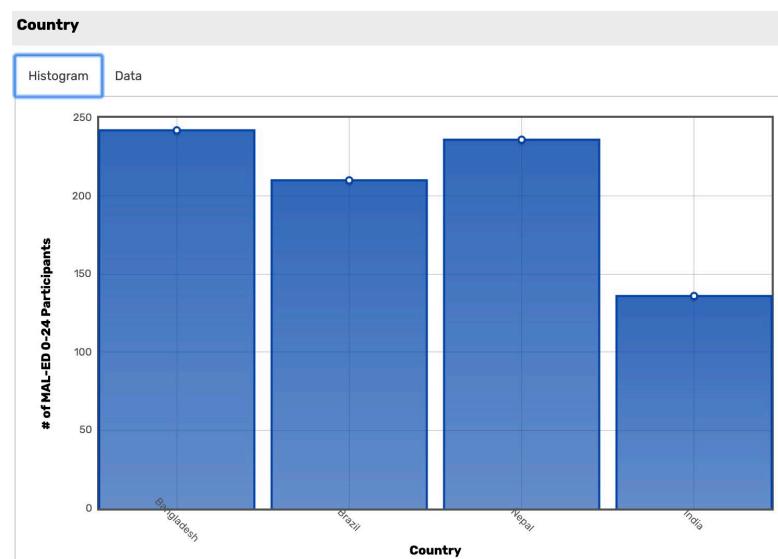
<sup>19</sup> After sorting the Results Table to arrange selected participants in descending order of “Total diarrheal episodes,” scroll down to determine that 4 of the selected participants had more than 20 diarrheal episodes.

31. Notice the histogram icons to the right of many of the column headers in the Results Table. Clicking on these histogram icons will open a popup window to visualize that variable's distribution. Which urban study site has the lowest number of participants who live in homes with improved sanitation?<sup>20</sup>

32. Can you determine the percentage of selected participants who had fewer than 2 diarrheal episodes over the course of the 2 year MAL-ED study?<sup>21</sup> Hint: After opening the relevant histogram popup window, click on the "Data" tab to determine the exact number of participants who had each possible number of diarrheal episodes observed for the "Total diarrheal episodes" variable.

---

<sup>20</sup> Click on the histogram icon next to the "Country" column header. Remember that we are looking at data from the 824 participants who met our selection criteria (urban study site, living in homes with improved sanitation). Of the 4 urban sites, India has the fewest participants who live in homes with improved sanitation.



<sup>21</sup> There were 264 participants (141 + 123) who had fewer than 2 diarrheal episodes in total. This is 32% of the 824 selected participants.

Total diarrheal episodes

Histogram Data

Search table

Rows per page: 100

Total diarrheal episodes	# of MAL-ED 0-24 Participants
0	141
1	123
2	88
3	94
4	62
5	56
6	57
7	47

## Exercise 4: Use the MAL-ED data on ClinEpiDB.org to answer the following questions – Independent work (45 minutes)

In this exercise, you will have an opportunity to explore the following questions on your own. If you have questions or get stuck, raise your hand and a moderator will come over to help you. Remember to clear any applied selection criteria before approaching each question. Try to start at the left side of the Search Wizard and move to the right as you apply filters. The “Find a filter” search box can help you easily find variables. Keep in mind that you can perform both Participant-level and Observation-level searches on the MAL-ED data, and think about whether the question is asking you to determine the number of participants or the number of observations before initiating a search.

### Independent work

#### Question 1: How does the number of Tanzanian participants with mothers who had at least 10 years of education compare to the number of South African participants?

1. Start from the home page. Navigate there using the ClinEpiDB logo located at the top of your screen. Determine if you need to perform a participant-level or an observation-level search and click on the appropriate icon on the MAL-ED study card to initiate the search.<sup>22</sup>
2. Navigate to the “Geographic region” box in the Search Wizard, and select “Tanzania.” Notice that you have selected 262 Participants.

The screenshot shows the ClinEpiDB search interface. At the top, a search wizard path is shown: "Geographic region" (highlighted with a red box) → "Households" → "Personal characteristics" → "Observations" → "Related observations" → "# Observations selected". Below this, a "Country" filter is applied, with "Tanzania" selected (highlighted with a red circle). The results table shows participant counts for various countries:

Country	Remaining Participants	Participants	Distribution	%
Bangladesh	265 (12%)	265 (12%)	██████	(100%)
Brazil	233 (11%)	233 (11%)	██████	(100%)
India	251 (12%)	251 (12%)	██████	(100%)
Nepal	240 (11%)	240 (11%)	██████	(100%)
Pakistan	277 (13%)	277 (13%)	██████	(100%)
Peru	303 (14%)	303 (14%)	██████	(100%)
Tanzania	314 (15%)	314 (15%)	██████	(100%)
<b>Tanzania</b>	<b>262 (12%)</b>	<b>262 (12%)</b>	<b>██████</b>	<b>(100%)</b>

<sup>22</sup> Initiate a participant-level search of the MAL-ED data.

#### Explore the Studies

The screenshot shows the ClinEpiDB homepage with four study cards:

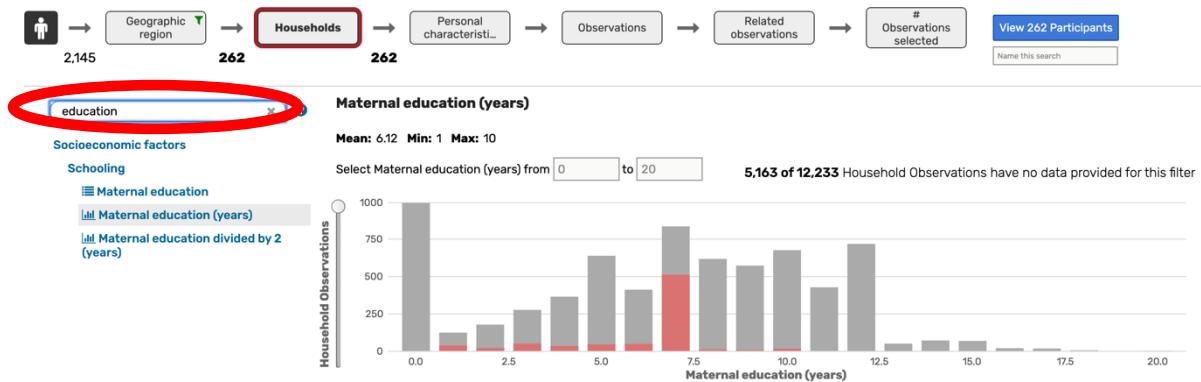
- GEMS1 Case Control**: Study Details. 7 S. Asian and African Sites, 2007-2011. Includes a brief description of the Global Enteric Multicenter Study (GEMS) and a link to download data.
- GEMS1A Case Control**: Study Details. 7 S. Asian and African Sites, 2011-2013. Includes a brief description of the GEMS1A study and a link to download data.
- India ICEMR Longitudinal Cohort Study**: Study Details. 2 sites in India from 2013-2015. Includes a brief description of the Center for the Study of Complex Malaria in India (CSCM) and a link to download data.
- MAL-ED Study**: Study Details. 8 Sites, 2009-2014. Includes a brief description of the Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development Study and a link to download data.

On the right side, there is a "News" section with three news items:

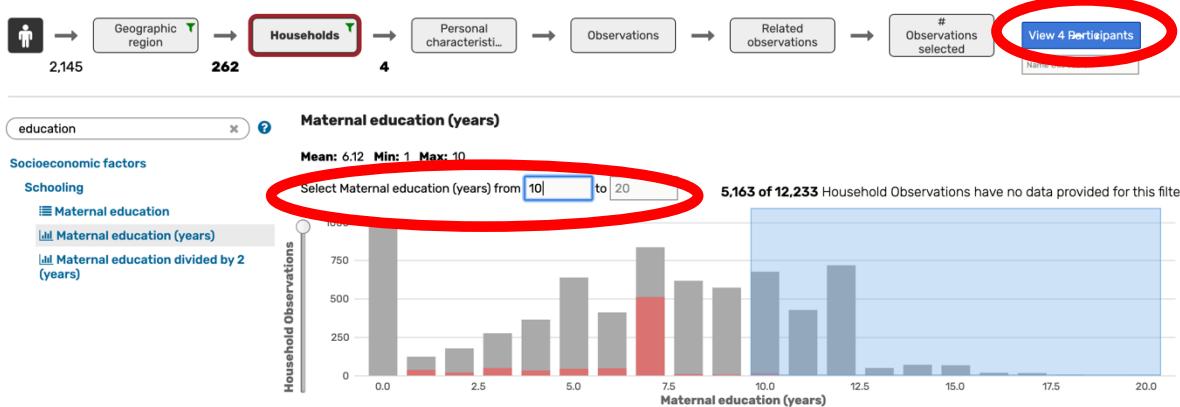
- ClinEpiDB 6 Released** (THU MAR 07 2019): We are pleased to announce the release of ClinEpiDB 6! ClinEpiDB in India March 13 - 17th, the ClinEpiDB outreach team will be attending a conference and conducting ... read more
- ClinEpiDB 5 Released** (SAT DEC 16 2018): We are pleased to announce the release of ClinEpiDB 5! New Features in this Release The analysis apps in the "Analyze Results" tab of the results table have now been updated... read more
- ClinEpiDB at ASTMH** (THU OCT 18 2018): ASTMH 2018 New Orleans The EuPathDB and See all news

A "Tweets by @ClinEpiDB" feed is also present at the bottom.

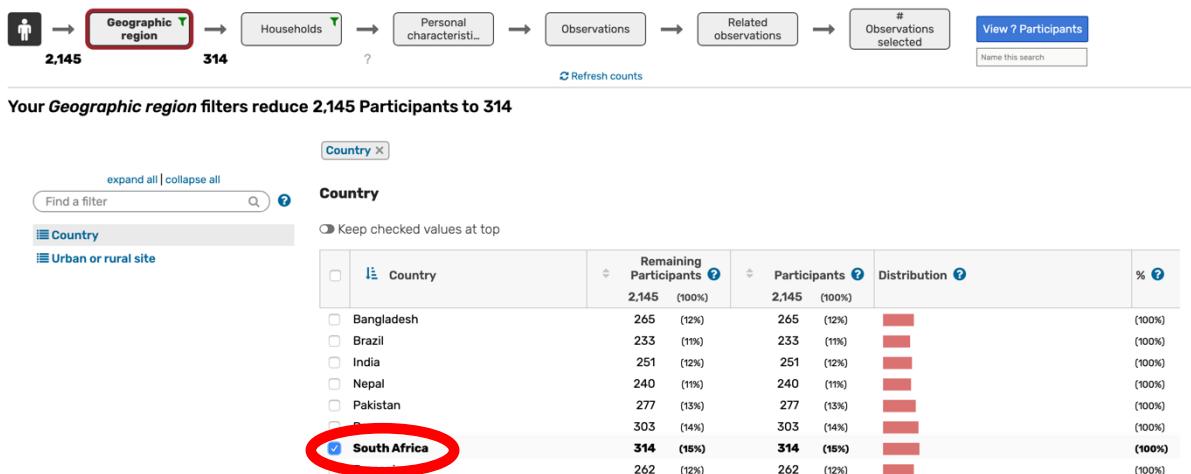
3. Now click on the “Household Observations” box in the Search Wizard. Use the “find a filter” search box to quickly navigate to variables related to education.



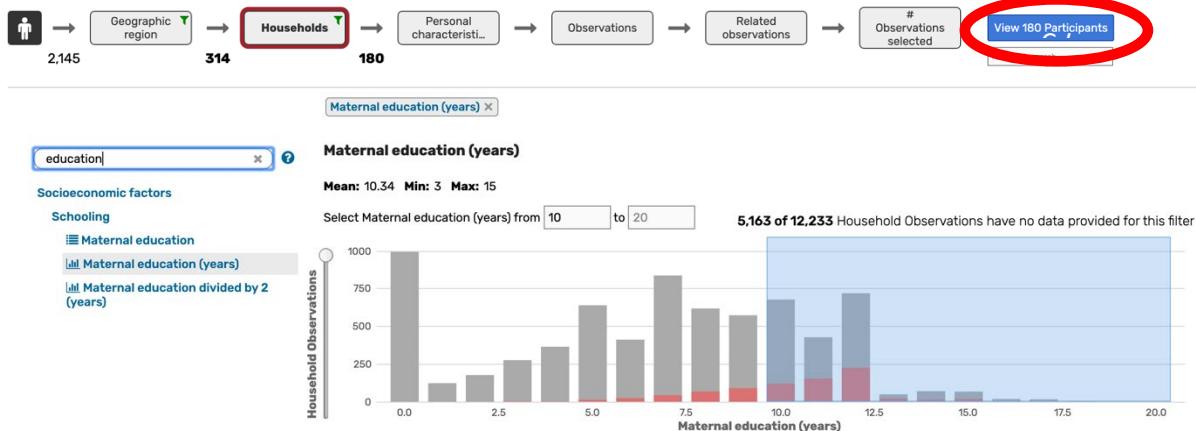
4. Click on the “Maternal education (years)” variable. Select observations with at least 10 years of maternal education. Note that there are 4 Tanzanian participants whose mothers had at least 10 years of education.



5. Without removing the filter for “Maternal education (years),” navigate back to the “Geographic region” box in the Search Wizard. Uncheck the box next to Tanzania and click to select South African participants. Notice that you now have 314 participants selected.

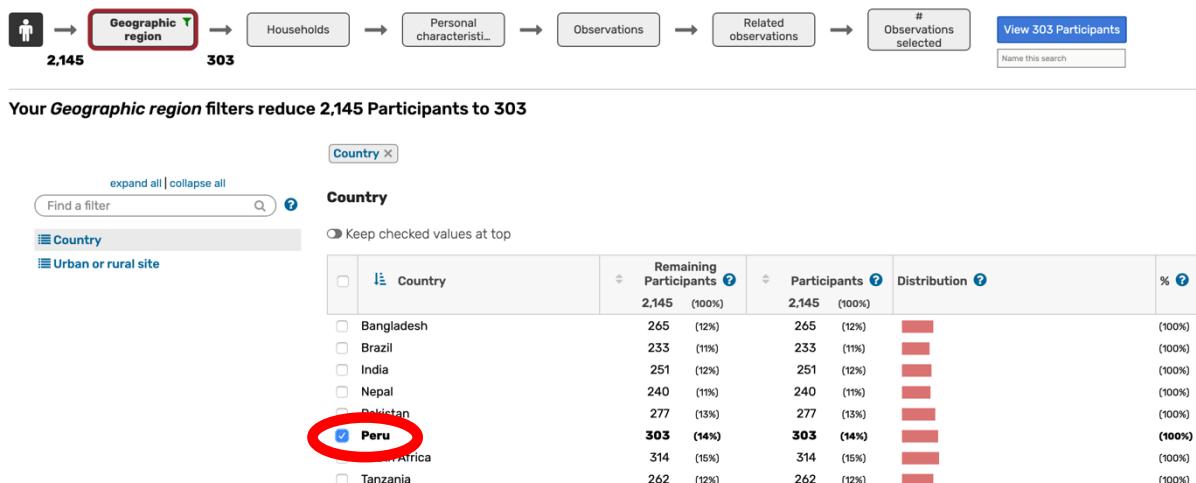


6. Click on the “Household Observations” box in the Search Wizard to see the number of participants whose mothers had at least 10 years of education in South Africa. “Maternal education (years)” should be the variable that automatically comes into view when you click on “Household observations,” but if not you can always use the “find a filter” search box to navigate it. Note that there are 180 South African participants whose mothers have at least 10 years of education.

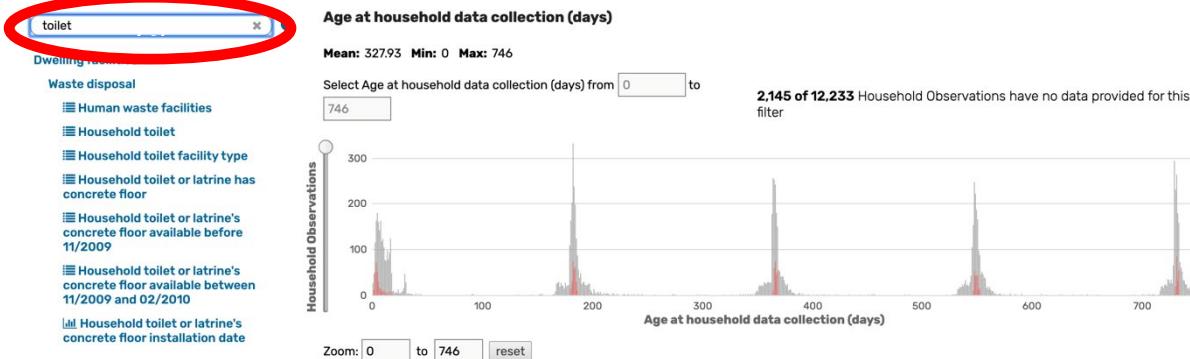


## Question 2: How many participants from Peru do not have any type of household toilet facility (in other words, houses that do not have any type of flush toilet or pit latrine)?

1. Start from the home page to clear your previous search results. Navigate there using the ClinEpiDB logo located at the top of your screen. Perform a participant-level search by clicking on the appropriate icon on the MAL-ED study card to initiate the search.
2. Navigate to the “Geographic region” box in the Search Wizard, and select “Peru.” Notice that you have selected 303 Participants.



3. Now click on the “Household Observations” box in the Search Wizard. Use the “find a filter” search box to quickly navigate to variables related to toilets.

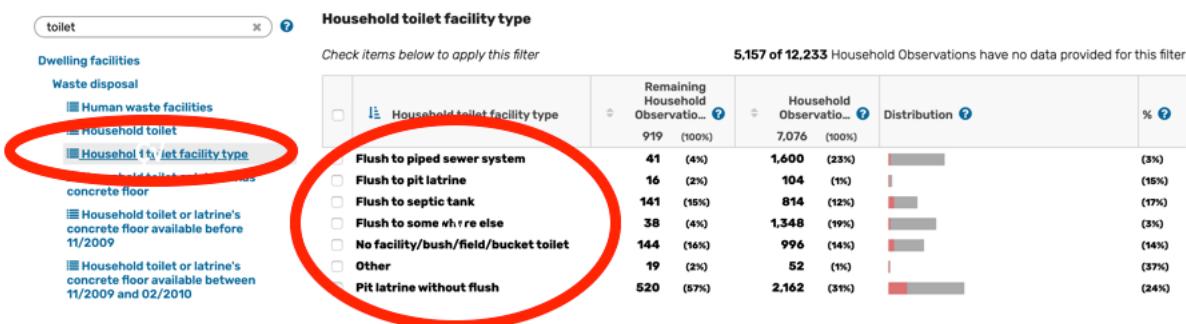


4. Click through the various variables related to household toilets to find the one containing data relevant to this question. “Household toilet” has data on whether or not the household had a toilet. Which variable contains the data you need to answer this question?<sup>23</sup>
5. Check the box that is relevant to households that do not have a household toilet or pit latrine. How many participants from Peru have ever lived in a house without a toilet or pit latrine during the course of the MAL-ED study?<sup>24</sup>

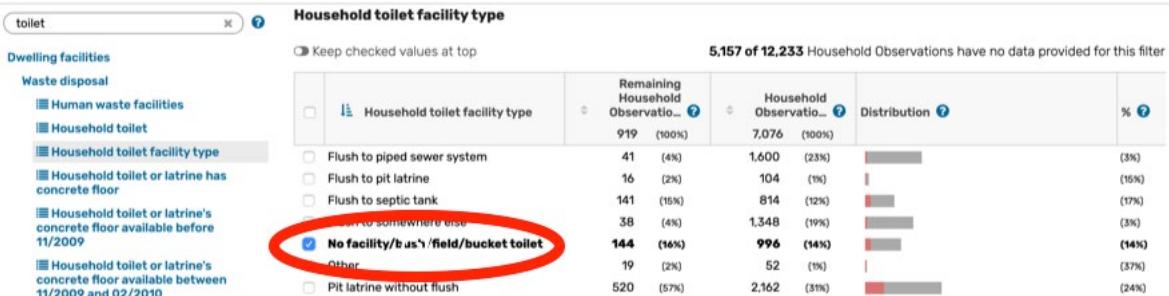
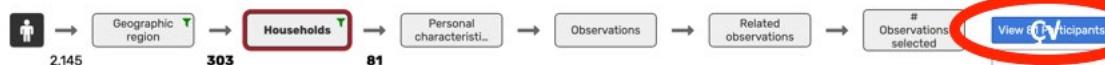
<sup>23</sup> Household toilet facility type contains a data on various types of flush toilets and pit latrines, including “No facility/bush/field/bucket toilet.”



No Households filters applied yet

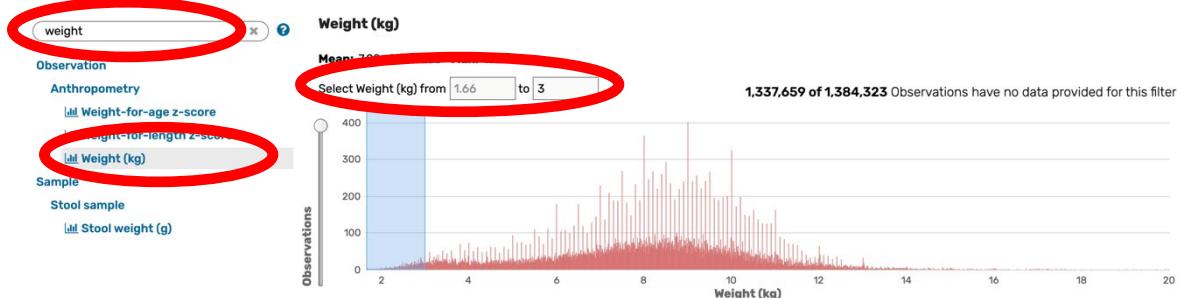


<sup>24</sup> 81 participants in Peru do not have any type of household toilet facility.

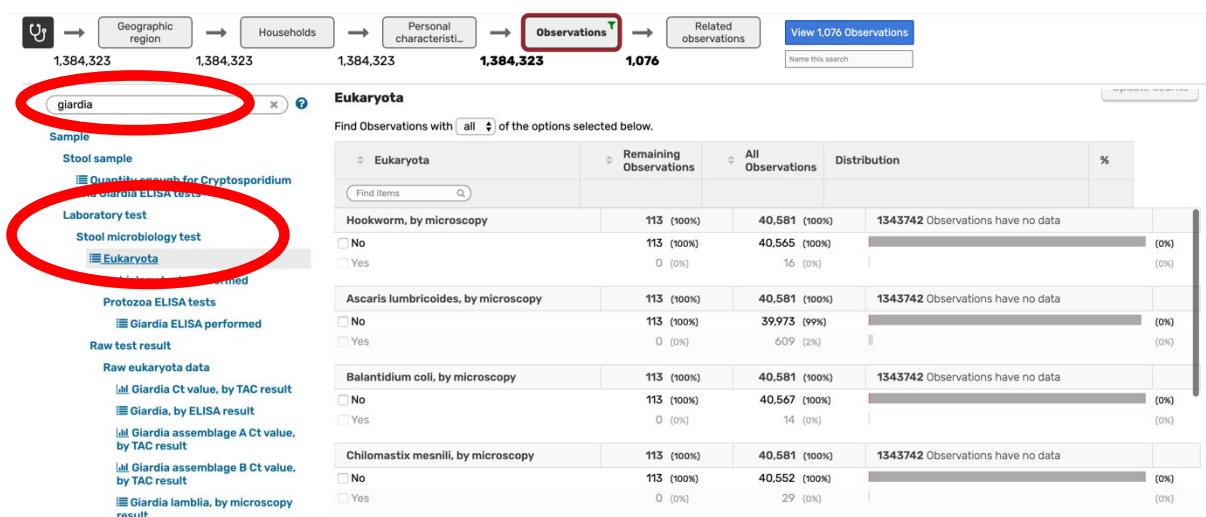


**Question 3: Is there a difference between the number of observation of *Giardia* in stool samples of participants who weighed 3kg or less compared to those who weighed 10kg or more? Which study site had the greatest number of *Giardia*-positive stool samples in participants who weighed at least 10kg? NOTE: this will be an observation-level search.**

1. Start from the home page to clear your previous search results. Navigate there using the ClinEpiDB logo located at the top of your screen. Perform an observation-level search by clicking on the appropriate icon on the MAL-ED study card to initiate the search.
2. Which box in the Search Wizard do you think you should click on to find variables related to participant weight? Why?<sup>25</sup> Click on this box in the Search Wizard.
3. Use the “find a filter” search box to navigate to variables related to the participants’ weight. Click on “Weight (kg)” in the variable hierarchy on the left side of the page to bring up a distribution of all of the weight observations during the MAL-ED study. Select all observations where the measured weight was 3kg or below. How many observations measured the participants’ weight to be 3kg or below?<sup>26</sup>



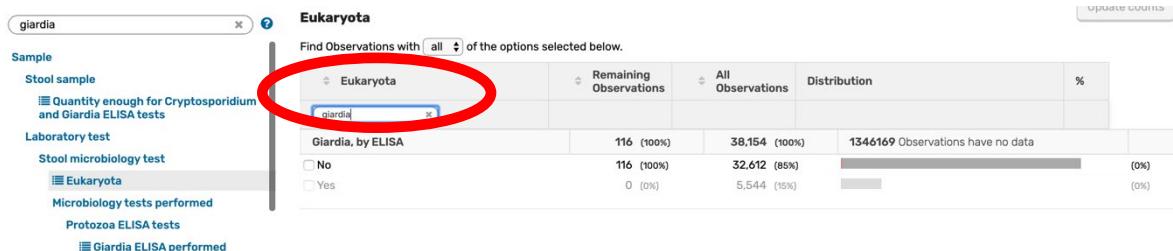
4. Next, use the “find a filter” search box to find variables related to *Giardia* detection. Notice that in the variable hierarchy, there is a “Eukaryota” variable under the “Stool microbiology test” sub-category. Click on “Eukaryota” to bring up processed microbiology data pertaining to the detection of various eukaryotic organisms.



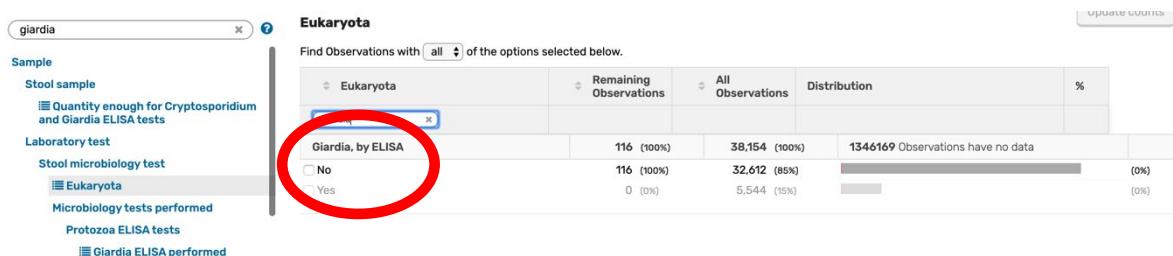
<sup>25</sup> Variables related to the participant’s weight are located under the “Observations” box in the Search Wizard. The “Observations” box generally contains variables that can change over time.

<sup>26</sup> There were 1,076 observations where participants’ weight was 3kg or below.

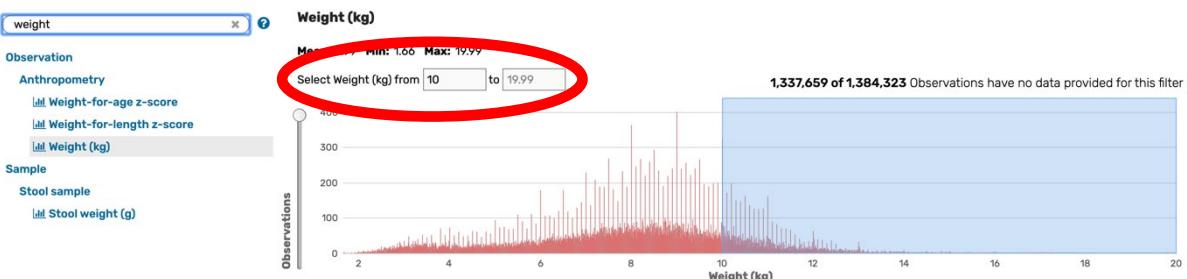
5. You can limit the processed microbiology data to data pertaining to the detection of “*Giardia*” under the “Eukaryota” variable by using the “Find items” search box.



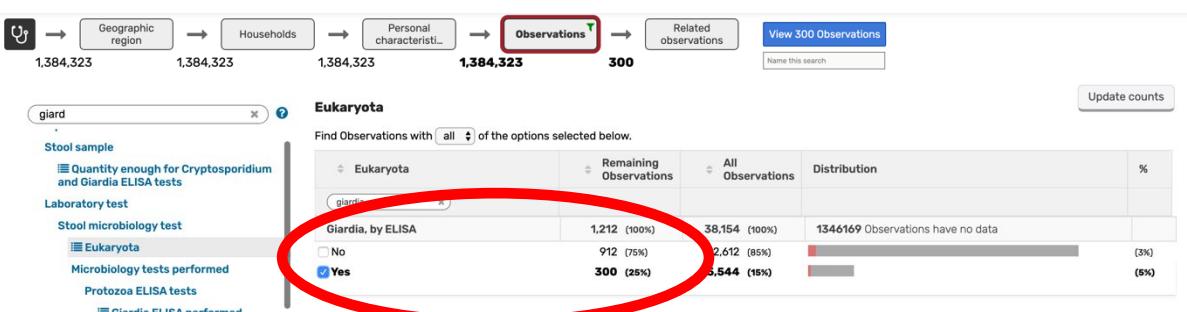
6. Notice that the option to select samples that were positive for “*Giardia*, by ELISA” is greyed out. In other words, you can check the box next to “No” for “*Giardia*, by ELISA” but you cannot check the box next to “Yes”. Why do you think this is?<sup>27</sup>



7. Navigate back to the “Weight (kg)” variable and select observations where the participant weighed more than 10kg.



8. Now look at “*Giardia*, by ELISA” under “Eukaryota.” Here, you can see that there are 300 observations in participants who weighed 10kg or more where the stool was *Giardia*-positive.



<sup>27</sup> Out of 38,154 samples tested for *Giardia* by ELISA, only 116 of these samples came from participants who weighed less than 3kg. Of these 116 samples, 100% are *Giardia*-negative by ELISA.

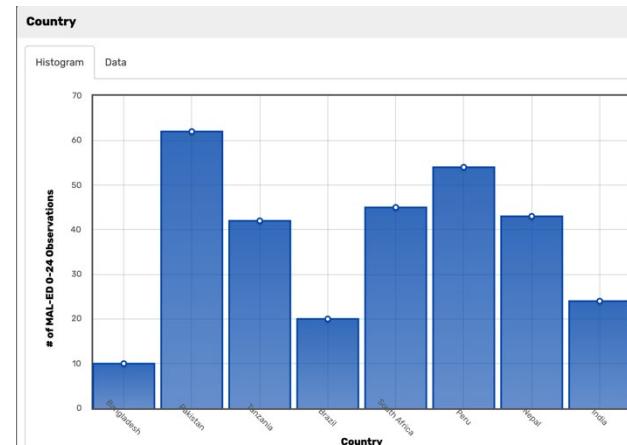
9. Next, the question asks us to determine which study site had the greatest number of stool samples collected from participants who weighed at least 10kg that tested positive for *Giardia*. Click on the blue “View Observations” box in the Search Wizard to bring up the results table pertaining to the relevant 300 *Giardia*-positive stool samples.



10. The results table contains one row for each observation that met the search criteria. See if your results table contains a column for “Country.” If not, click on the “Add Columns” button to add this variable to the results table. Click on the histogram icon next to “Country” in the column header of the results table to bring up the distribution which country each of these 300 stool samples were collected from.

Observation Id	Participant ID	Sex	Age (days)	Day of ALRI episode	Day of diarrheal episode (ex LM induced)	Exclusively breastfed
253abeac25cf3daa	fc7c2bd836e7f9e2	Male	Peru	609	N/A	No
d75d0ef9554f8ed	fc585ab71a65f251	Male	Brazil	367	N/A	No
1d1e7b66eb7867b	f97227e56c363c2e	Female	Peru	580	N/A	No
f40ae9be69e28e6a	f9389f2d265de4b	Male	Pakistan	729	N/A	No
62c4bde59255ca8b	f90e6cb4fe80a88a	Male	Peru	732	N/A	N/A
3ca9f7e130f5f6a8	f8b943f8344fc42	Female	Peru	641	N/A	No
4d64b9ccce357a39	f8893cd7e26146b	Male	Pakistan	547	N/A	No
61a90805b00d13e7	f800db9b03848a39	Male	Pakistan	547	N/A	No
0ec0ce2059184a8b	f7340ffa8915615e	Female	South Africa	729	N/A	N/A
5346d12467dee920	f625eb43ff2564c	Male	India	730	N/A	No
44c66c8f83dd52c	f620beb7831604e9	Female	Tanzania	547	N/A	No

11. From this histogram, it is apparent that the greatest number of *Giardia*-positive stool samples collected from participants weighing at least 10kg came from Pakistan. Can you determine the exact number of stool samples meeting the search criteria that came from Pakistan?<sup>28</sup>



<sup>28</sup> 62 of the 300 samples came from participants at the Pakistan study site. To determine this, hover over the bar corresponding to Pakistan. Alternatively, you can click on the “Data” tab to bring up the data in tabular format.

Country	# of MAL-ED 0-24 Observations
Bangladesh	10
Pakistan	62
Tanzania	42
Brazil	20
South Africa	45
Peru	54
Nepal	43
India	24

**Question 4: Compare the average number of diarrheal episodes in the first year of life in participants from Pakistan versus Bangladesh for participants that were followed until at least 1 year of age.**

1. Start this exercise from the home page. Navigate there using the ClinEpiDB logo located at the top of your screen. For this exercise, we will be conducting an observation-level search of the MAL-ED data. Click on the “Observations” icon on the MAL-ED Study Card to get started.
2. Use the search wizard to navigate to “Geographic region”, and select participants from the Pakistan study site. Note that there are 190,280 observations of the 277 subjects from Pakistan.



Your **Geographic region** filters reduce 1,384,323 Observations to 190,280

**Country**

expand all | collapse all

Find a filter

**Country**

**Urban or rural site**

Keep checked values at top

<input type="checkbox"/> Country	Remaining Participant <small>?</small>	Participant <small>?</small>	Distribution <small>?</small>	% <small>?</small>
<input type="checkbox"/> Bangladesh	265 (12%)	265 (12%)		(100%)
<input type="checkbox"/> Brazil	233 (11%)	233 (11%)		(100%)
<input type="checkbox"/> India	251 (12%)	251 (12%)		(100%)
<input type="checkbox"/> Nepal	240 (11%)	240 (11%)		(100%)
<input checked="" type="checkbox"/> <b>Pakistan</b>	<b>277 (13%)</b>	<b>277 (13%)</b>		<b>(100%)</b>
<input type="checkbox"/> South Africa	303 (14%)	303 (14%)		(100%)
<input type="checkbox"/> Tanzania	314 (15%)	314 (15%)		(100%)
	262 (12%)	262 (12%)		(100%)

3. Next, navigate to “Observations” box in the Search Wizard. Restrict the selected observations to those that occurred within the first year of life by selecting “Age (days)” from 0 to 365. Note that the number of observations is further reduced to 97,158.

Your **Observations** filters reduce 190,280 Observations to 97,158

**Age (days)**

expand all | collapse all

Find a filter

**Observation**

- Age (days)**
- Diarrhea, ALRI, fever, or vomiting present
- Observation details
- Anthropometry
- Diarrheal episodes
  - Diarrheal data present
  - Study-defined diarrhea
  - 1st day of diarrheal episode
  - Day of diarrheal episode (ex LM included)

Mean: 360.51

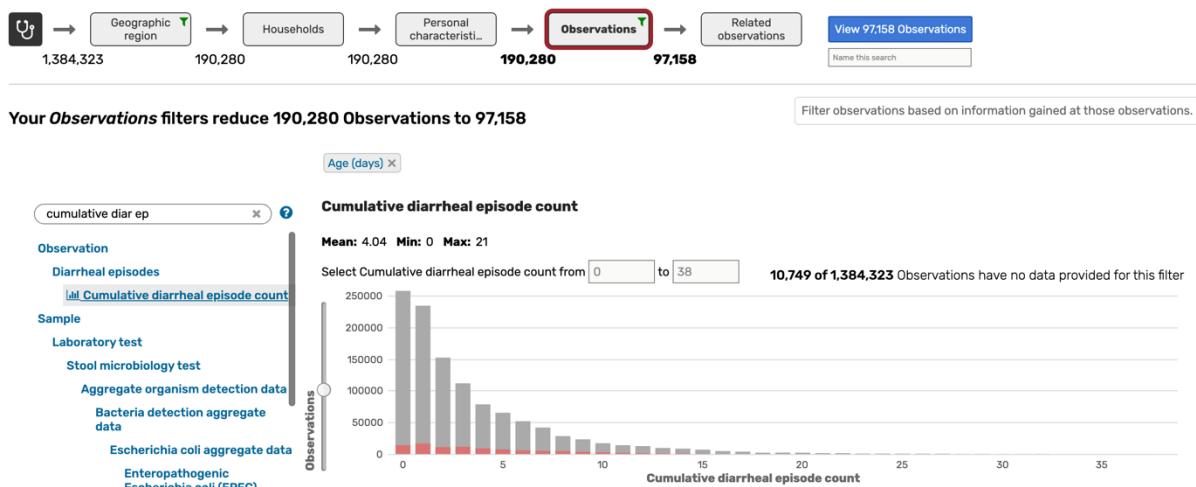
Select Age (days) from  to

Observations

Age (days)

4. We will want to look at the data for “Cumulative diarrheal episode count,” so navigate to that variable now. Let’s think about how to approach this question.

- The Search Wizard indicates that, when we filter the data to observations taken within the first year of life, there are 97,158 observations from 277 participants (~351/participant on average). This makes sense because illness surveillance was conducted on a daily basis. Where can you find the information that illness surveillance was conducted on a daily basis?<sup>29</sup>
- The “Cumulative diarrheal episode count” is an increasing count of all diarrheal episodes a participant had since beginning follow-up. Thus, “Cumulative diarrheal episode count” was calculated for each day of illness surveillance. We currently have 97,158 observations of the “Cumulative diarrheal episode count” selected.
- We are going to have to restrict our criteria for selecting observations further. How would you do this?<sup>30</sup>

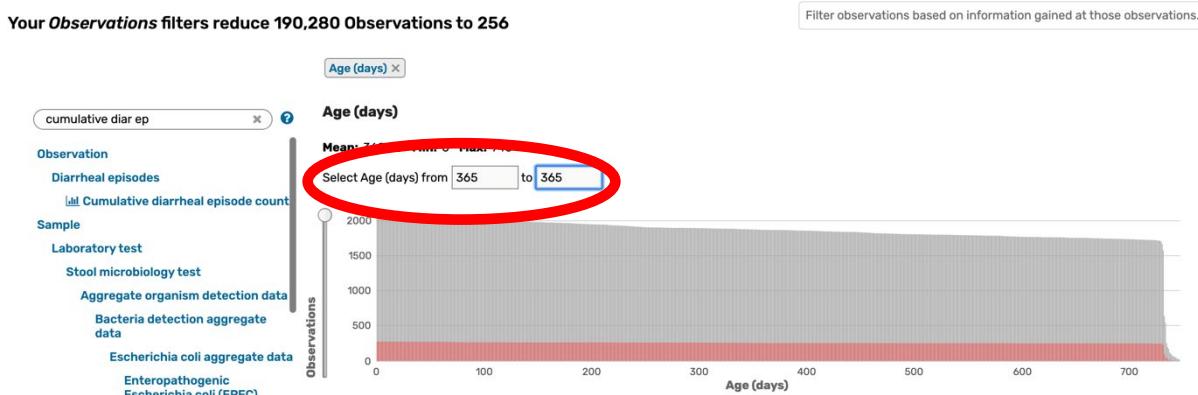


5. Navigate back to the “Age (days)” variable. We can restrict our observations to those that were collected at exactly 365 days of age by selecting “Age (days)” from 365 to 365 days. Note that we now have 256 observations from 277 participants. Why don’t we have 277 observations?<sup>31</sup>

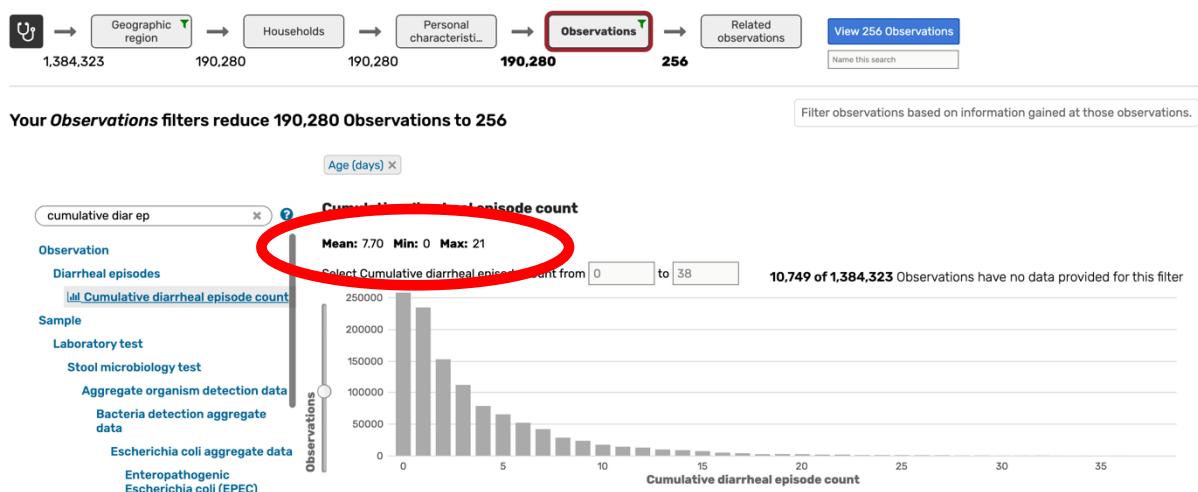
<sup>29</sup> You can find this type of information on the Data Set page for the MAL-ED study.

<sup>30</sup> We want to determine the cumulative number of diarrheal episodes that each participant had by 365 days of age. We are not interested in the cumulative number of diarrheal episodes that a participant had on 0-364 days of age, so we will need to filter out observations that occurred on these days.

<sup>31</sup> Some participants dropped out of the study before reaching 365 days of age.



6. Now navigate the “Cumulative diarrheal episode count” and determine the mean number of diarrheal episodes that participants from Pakistan had by 1 year of age. Note that, based on the selection criteria that we applied upstream, only participants who have been followed for at least 365 days will be included when calculating this mean. Participants from Pakistan averaged 7.7 episodes of diarrhea by 1 year of age.



7. Repeat this procedure for the Bangladesh site. What was the mean number of diarrheal episodes that participants from Bangladesh had by 1 year of age?<sup>32</sup> Hint: you can navigate back to the “Country” variable without removing any of the downstream selection criteria that you have applied.

<sup>32</sup> 232 participants from Bangladesh were followed until at least 1 year of age. These participants had an average of 3.69 episodes of diarrhea by 1 year of age.