What is the prevalence of wasting and stunting in the GEMS and GEMS1A data sets and how do these relate to the presence of different pathogens? (30 min)

To explore how anthropometric data is distributed in the GEMS populations, this exercise will take you extensively through the query wizard. You will discover how you can explore using the data filters and their corresponding histograms and bar graphs.

The GEMS1 and GEMS1A measured height and weight at two time points for most children. Cases were measured at presentation to the clinic and if given rehydration therapy, were weighed during this process and then again at 60-day follow-up. Controls were measured at enrollment and follow-up. While this study is case/control without significant longitudinal follow-up we may be interested in what the prevalence of wasting and stunting are in the population at enrollment compared to 60-day follow-up and in different sub-groups of data.

1. Navigate to http://clinepidb.org in your browser.

For GEMS, "Participants" refers to 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 at this time mothers, fathers, or caregivers are not considered participants so the data is only available as associated with the child.

You can access the data about these participants in a few ways:

1. Through the navigation bar at the top of the page under Search a Study>GEMS1 Case Control>Search Participants
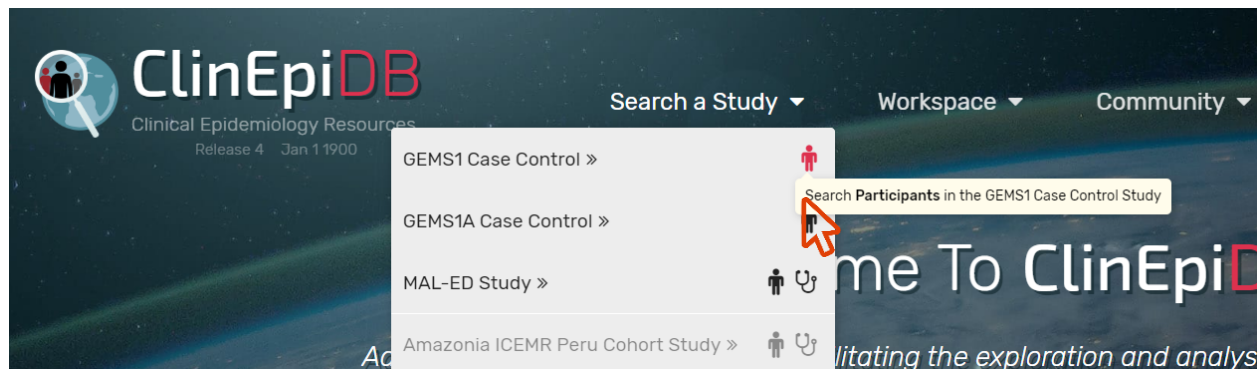
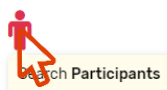2. Click on the 'Participant' icon on the GEMS1 card:
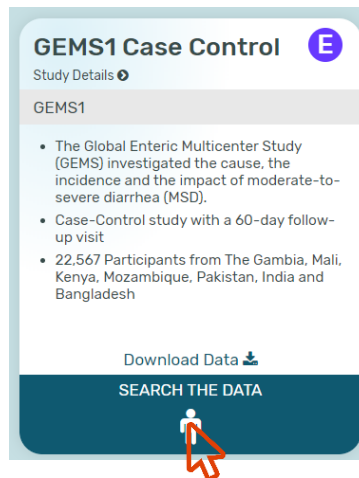
3. Or click on the main text of the GEMS1 Case Control Card and a new page with information about GEMS1 will load. On this page, find 'SEARCH THE DATA', click on the small human icon next to this.

Any of these three options will take you to the GEMS1 search page which includes a Search wizard.

What is the prevalence of stunting of GEMS1 participants at Enrollment?

The purpose of the search wizard is two-fold. First, it creates a simple way to categorize components of the data allowing for a step-wise approach to building searches. Second, it allows you to explore the data to see what the raw number and distribution of characteristics are in both the full dataset or filtered data. Spend some time reading and mouse over the different search wizard boxes.

Select a Set of Participants (GEMS)

Choose Case/Control → Geographic region → Households → Personal characteristi... → Observations → Related Case/Control → Results Options

View 22,567 Participants

ℹ Learn about the GEMS1 Case Control

22,567   22,567

No *Choose Case/Control* filters applied yet

Both Cases and Controls ▼

Choose whether to query just cases, just controls or both.

Notice that there is a number below the black square Participant icon. This number represents the total number of Participants that are included in this dataset.

Finding Filters and Examining Data in the Filters

1. Notice the first open filter listed is 'Choose Cases/Control' as a specific selection at the top where you can restrict your search to either cases, controls or the entire dataset. By default, we will search the entire dataset. Leave it this way for now.

2. Without applying a geographic region filter, can you tell which region had the most number of participants?[1] Notice that you can sort the columns of data by clicking the small up and down arrows in the heading of the columns.

Country

Check items below to apply this filter

	Country		Remaining Participants ❓		Participants ❓	Distribution ❓	% ❓
			22,567 (100%)		22,567 (100%)		
	Bangladesh		3,859 (17%)		3,859 (17%)		(100%)
	India		3,582 (16%)		3,582 (16%)		(100%)
	Kenya		3,359 (15%)		3,359 (15%)		(100%)
	Mali		4,097 (18%)		4,097 (18%)		(100%)
	Mozambique		1,976 (9%)		1,976 (9%)		(100%)
	Pakistan		3,096 (14%)		3,096 (14%)		(100%)
	The Gambia		2,598 (12%)		2,598 (12%)		(100%)

3. Next, click on the Personal characteristics filter box. This will reveal a slightly more complex filtering table that includes major categories in blue text on the left and summary information on the right. Variables included as filters in this section generally are only reported once per participant and can also relate to information about how the study was conducted (Administrative Information for example)

---

[1] Mali

4. Click on the 'Observations' category to find information on variables that were measured during the study. By default, the 'Observation Type' variable is open, select 'Enrollment' from the values.

5. To estimate the prevalence of stunting in both cases and controls we first have to find the height-for-age z-score variable. We can do this by typing in the 'Find a filter' box on the left'. If you start typing it will narrow your selection.

6. We know that data can be either categorical or numerical. For numerical data in our database that has greater than 10 values, the data is displayed as a histogram rather that in table format. This allows you to see the distribution of values of your selected participants. Basic statistics about the selected data is displayed immediately above in the histogram summary information. This can easily be determined by the statistics displayed above the histogram. This is calculated based on any filters you've applied previously. How many participants had an HAZ less than or equal to -2.00 [Hint: You can select the HAZ range using by highlighting a selected area or by entering it in the boxes above.]
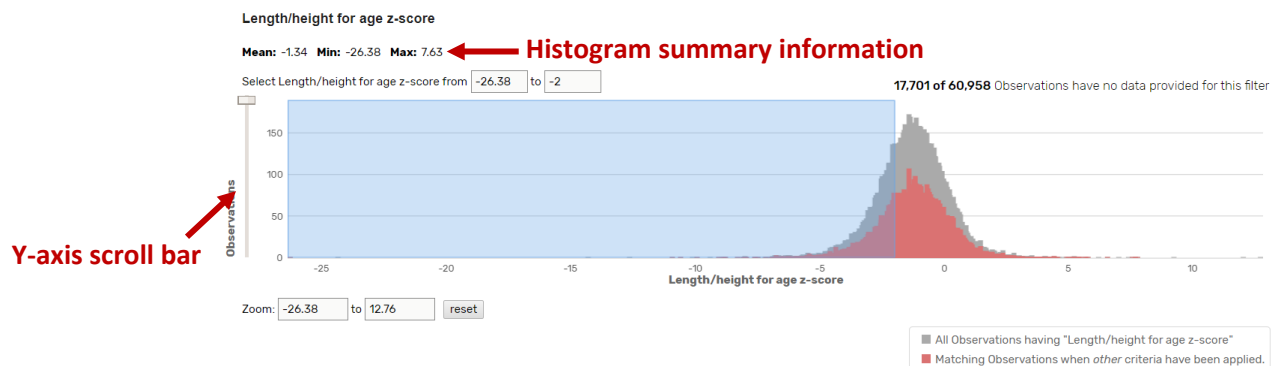
7. How many participants remain after you make this selection? You can easily convert this to a proportion by dividing the number of participants by the overall sample size (22,567). This indicates the prevalence of stunting overall in all cases and controls. To see how this differs in Cases versus Controls….you can navigate back to the 'Choose Case/Control' search wizard step

and see how your selections update the participant counts. Of 9,439 cases 2,660 (28.1%) had an HAZ score less than or equal to -2. Of 13,128 controls, 3,778 (28.7%) were stunted.

8. Next, determine the prevalence of wasting (weight-for-height) overall at enrollment and in cases and controls.

   Enrollment

       a. WHZ <=-2 Overall = _____

       b. WHZ <=-2 Cases   = _____

       c. WHZ <-2 Controls = _____

   60-day follow-up

       d. WHZ <=-2 Overall = _____

       e. WHZ <=-2 Cases   = _____

       f. WHZ <-2 Controls = _____

9. Examine all available filter categories in blue text on the left-hand side of the page in the Observations section of the search wizard. Click on the 'expand all' link at the top of the list, above the "Find a filter" search box. This will reveal all the types of data in each of the subcategories. Scroll down and read through these filters. Spend some time clicking through the filter and examining the distribution of participant data in each filter.
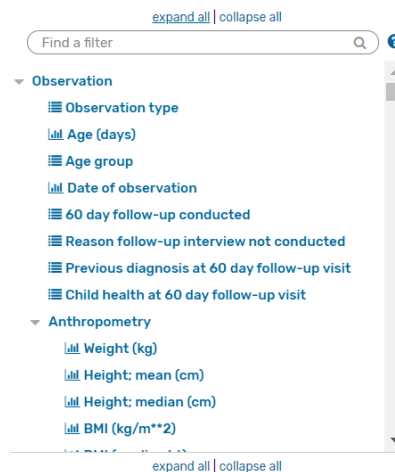
10. Can you find the information about Shigella? How many cases and controls had Shigella detected by traditional methods? What about by TAC?

Hints: Remember you can clear or view any applied filters as you explore the data using the green filter icon at the top of the search wizard or in any of the search wizard category boxes.