

Interactive II: Complex Epidemiology Queries

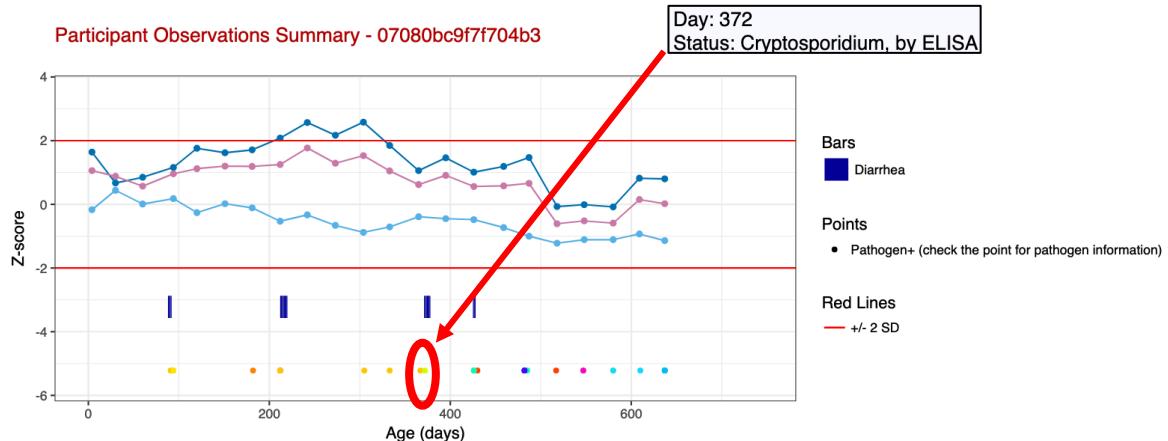
Related Observations

Exercise 1: Building a Search with Related Observations

In this exercise, we will go through an example together to perform a more advanced search using the “Related Observations” function. Several studies on ClinEpiDB included extensive collection of longitudinal information. The ability to search these data easily is extremely valuable. Integrating the “Related Observations” function into your searches allows you to examine data over time.

Before we start, consider the graph below. This graph plots height-for-age z-scores (light blue), weight-for-height z-scores (dark blue), and weight-for-age z-scores (pink), measured over time, for one of the participants in the MAL-ED study.

Z-score Observations Graph



During follow-up, this participant had four episodes of diarrhea (the blue hash bars on the graph indicate diarrheal episodes). The thickness of each blue hash bar is proportional to the duration that diarrheal episode lasted. There are two diarrheal episodes that appear to be greater than a few days in length.

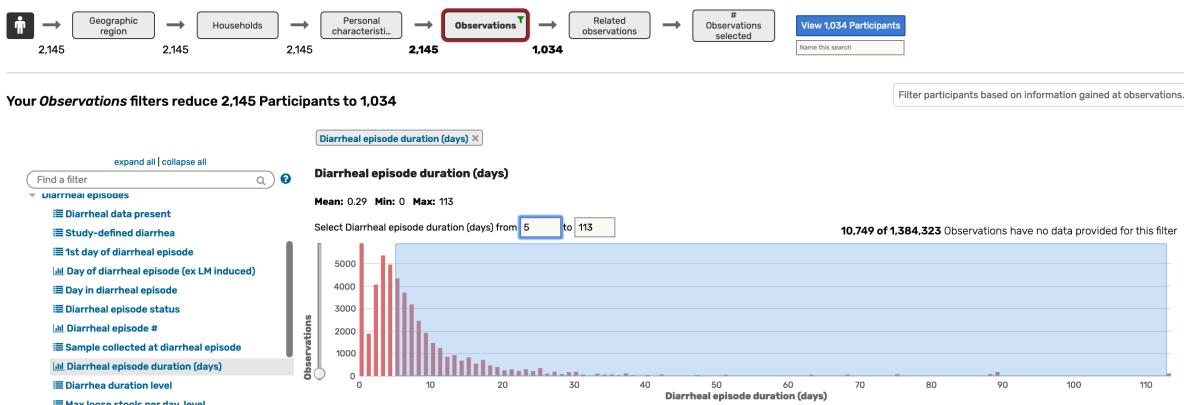
You can also see that this participant had multiple instances of positive laboratory test results for microbiology, indicated by the colored dots at the bottom of the graph (different colored dots represent the detection of different pathogens). One question you might want to ask is, **“Did participants tend to test positive for a certain type of pathogen before an episode of diarrhea that was five or more days in duration?”**

To explore this question, your first step might be to explore the data and see how many participants tested positive for Cryptosporidium within the 14 days before a diarrheal episode that was five or more days in duration. Then you might want to repeat the same question for any Campylobacter.

You can do this by using the Related Observations functionality! We will start with a very simple related observations query. For this exercise, to start, we will modify as few filters as possible to keep things straightforward.

Question 1: How many participants in the MAL-ED study tested positive for *Cryptosporidium* within the 14 days before a diarrheal episode that was five or more days in duration?

1. Start a participant-level search of the MAL-ED data. To answer this question, first select participants who had a diarrheal episode of at least 5 days in length at any point during the study.
 - Navigate to the “Observations” box in the Search Wizard and click on the “Diarrheal episode duration (days)” filter.
 - Next, select observations that had a duration of longer than 5 days using the highlighted portion of the histogram or using the range selection boxes above the graph.



- Notice that the number of participants has been reduced to 1,034. This means that 1,034 participants had at least one diarrheal episode of at least 5 days of duration at any point in the study. Keep in mind that a participant could have had more than one episode of diarrhea that is at least 5 days of duration.
2. Now, click on the “Related Observations” box in the Search Wizard. “Related Observations” are used to narrow the dataset based on the relationship between two variables at two different points in time.



3. Click on the small check box next to: “Enable the advanced Related observations filter below. It allows you to restrict Observations by relating them to your choice of Related observations.”

Enable the advanced Related observations filter below. It allows you to restrict Observations by relating them to your choice of Related observations.

After you have enabled the Related Observations filter, notice that the variables are now accessible. Take a look at the various variables and categories that appear in the variable hierarchy on the left side of the page. Do you notice anything familiar about these variables?¹

¹ The variables and categories available on the Related Observations step in the Search Wizard are duplicates of those that were available in the Observations step in the Search Wizard. The Related Observations step allows us to take the timing of different episodes, results, or data collection into account so we can select for data that occurred before (or after) something else had occurred.

4. To answer this question, we need to identify participants who tested positive for *Cryptosporidium* within the 2 weeks **BEFORE** a diarrheal episode that lasted for at least 5 days. In the Observations step in the Search Wizard, we previously selected participants who had diarrheal episodes that lasted for 5+ days. We now want to narrow down the selected participants further, and keep only the selected participants who also had a positive lab test for *Cryptosporidium*. However, we don't want to keep all selected participants who **EVER** tested positive for *Cryptosporidium* over the 2 years of follow-up. The timing of the *Cryptosporidium* test in relation to particular diarrheal episodes is important.

How would you complete the 4 boxes in the sentence below to properly account for the timing of *Cryptosporidium*-positive laboratory results in relation to diarrheal episodes lasting at least 5 days?² *HINT: In the sentence, "Observations" refers to the observations we selected in the previous step (diarrheal episodes that lasted at least 5 days). "Related observations" refers to the observations we are selecting now (positive *Cryptosporidium* test results).*

Observations that are to days the Related observations specified below

5. Now we will want to specify that we are selecting *Cryptosporidium*-positive observations for our Related Observation. Use the "find a filter" search box to find variables related to *Cryptosporidium* detection. Remember that there is a "Eukaryota" variable under the "Stool microbiology test" sub-category, which contains processed microbiology data pertaining to the detection of various eukaryotic organisms. Click the "Yes" box next to "*Cryptosporidium*, by ELISA" in the pathogen selection table under the "Eukaryota" category.
- How many participants in the MAL-ED study tested positive for *Cryptosporidium* within the 14 days before a diarrheal episode that was five or more days in duration?³

Find items	Remaining Related Observations	All Related Observations	Distribution	%
Eukaryota	40,552 (100%)	40,552 (100%)	40,552 (100%)	(100%)
No	29 (0%)	29 (0%)	29 (0%)	(100%)
Yes	38,147 (100%)	38,147 (100%)	1346176 Related Observations have no data	(100%)
Cryptosporidium, by ELISA	36,396 (95%)	36,396 (95%)	36,396 (95%)	(100%)
No	1,751 (5%)	1,751 (5%)	1,751 (5%)	(100%)
Yes	38,617 (100%)	38,617 (100%)	1345706 Related Observations have no data	(100%)
Cyclospora, by microscopy	38,600 (100%)	38,600 (100%)	38,600 (100%)	(100%)
No	17 (0%)	17 (0%)	17 (0%)	(100%)
Yes	38,583 (100%)	38,583 (100%)	1343742 Related Observations have no data	(100%)
Endolimax nana, by microscopy	40,581 (100%)	40,581 (100%)	40,581 (100%)	(100%)
No	40,492 (100%)	40,492 (100%)	40,492 (100%)	(100%)
Yes	89 (0%)	89 (0%)	89 (0%)	(100%)
Entamoeba coli, by microscopy	40,581 (100%)	40,581 (100%)	40,581 (100%)	(100%)
No	40,427 (100%)	40,427 (100%)	40,427 (100%)	(100%)
Yes	154 (0%)	154 (0%)	154 (0%)	(100%)
Entamoeba histolytica, by ELISA	38,069 (100%)	38,069 (100%)	38,069 (100%)	(100%)
No	37,847 (99%)	37,847 (99%)	37,847 (99%)	(100%)
Yes	223 (1%)	223 (1%)	223 (1%)	(100%)

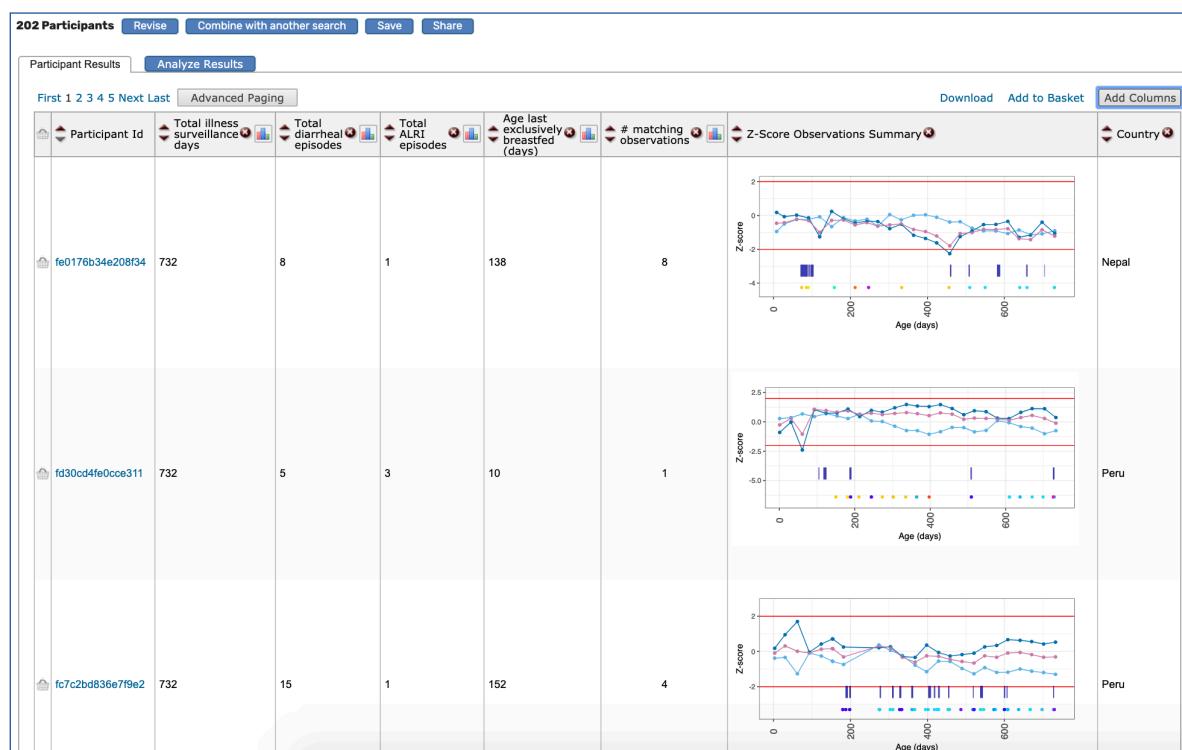
2

Observations that are to days the Related observations specified below

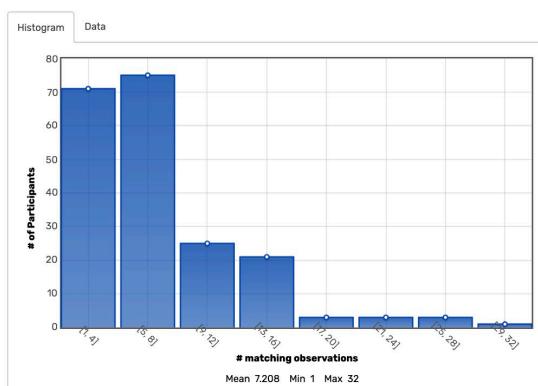
³ Using the "Related Observations" filter, the remaining participants was reduced from 1,034 participants who had a diarrheal episode at least 5 days in length to 202 participants who had a 5+ day diarrheal episode preceded by a positive *Cryptosporidium* result within the previous 2 weeks.

(Optional) Question 1b: Out of the participants in the MAL-ED study who tested positive for *Cryptosporidium* within the 14 days before a diarrheal episode that was five or more days in duration, how many participants met the search criteria exactly 1 time?

1. Click on the blue “View 202 Participants” box in the Search Wizard to see the “Results Table” for the participants who tested positive for *Cryptosporidium* within the 14 days before a diarrheal episode that lasted for at least 5 days.
2. If your default columns in the Results Table are different from those shown in the screenshot below, click on the “Add Columns” button located at the top right corner of your “Participant Results” tab and add the desired columns.



3. “# of matching observations” indicates the number of times the participant met the specified search criteria. Click on the graph icon in the column header to open the histogram.



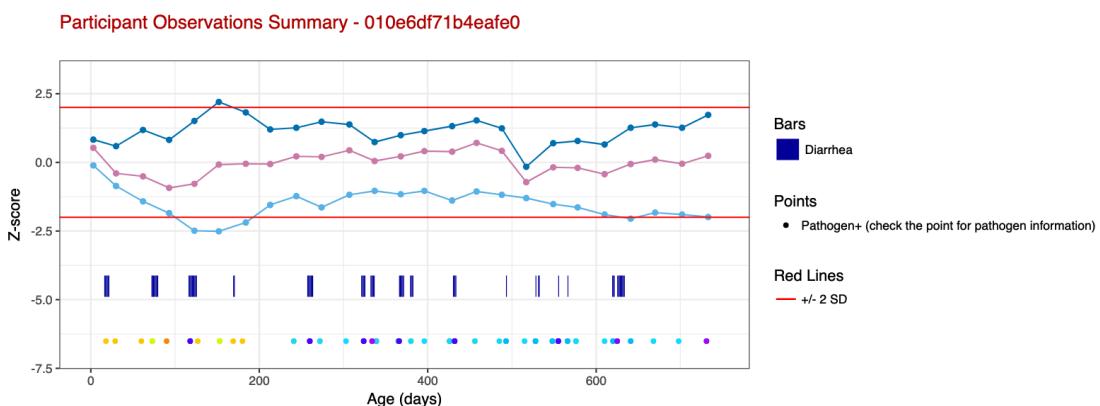
- Some participants met the search criteria more than once. During the course of the study, they had at least 2 episodes of diarrhea that were 5+ days in duration preceded by a positive *Cryptosporidium* test within the preceding 2 weeks.
- How many participants met the search criteria exactly 1 time?⁴ Hint: Click on the “Histogram” icon next to “# matching observations” and navigate to the “Data” tab.

⁴ 16 participants met the search criteria exactly 1 time.

(Optional) Question 1c: How old was MAL-ED participant 010e6df71b4eafe when they had a positive *Cryptosporidium* test result that preceded a diarrheal episode of at least 5 days?

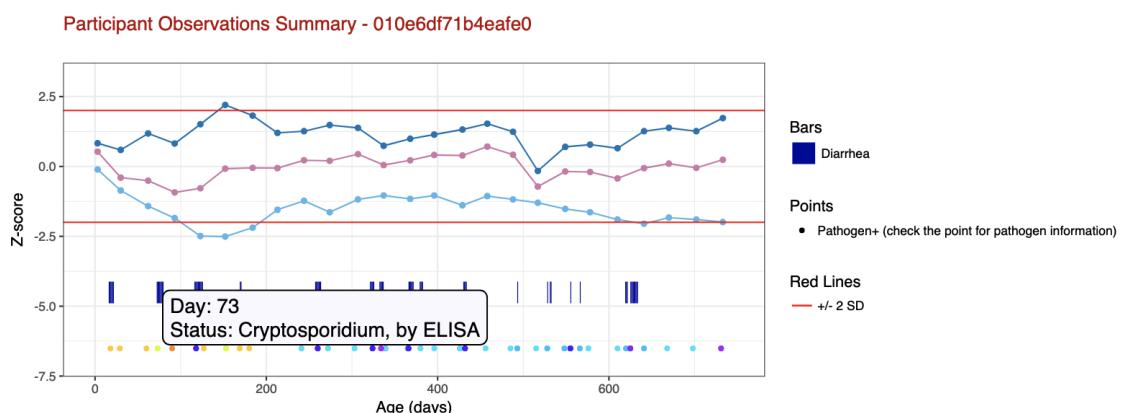
1. Sort the Results Table by Participant ID by clicking on the “Participant ID” column header until Participant 010e6df71b4eafe appears in the first row.
2. Click on 010e6df71b4eafe to bring up data on this specific individual.
3. In the “Z-score Observations Graph”, notice the colored dots on the bottom of the graph. These colored dots represent positive microbiology test results for different organisms. Mouse over these dots to see information about each positive microbiology test result. Can you determine how old Participant 010e6df71b4eafe was when they had a positive *Cryptosporidium* test result that preceded a diarrheal episode of at least 5 days using this figure?⁵

▼ Z-score Observations Graph



⁵ Participant 010e6df71b4eafe had a positive *Cryptosporidium* test result preceding a diarrheal episode of at least 5 days at 73 days of age. To determine this, hover over the dots representing positive microbiology test results to determine the age at testing and the organism detected. On day 73, this participant had a positive *Cryptosporidium* test. Looking at the tick marks representing diarrheal episodes, this participant had a longer episode of diarrhea soon after this *Cryptosporidium* detection.

▼ Z-score Observations Graph



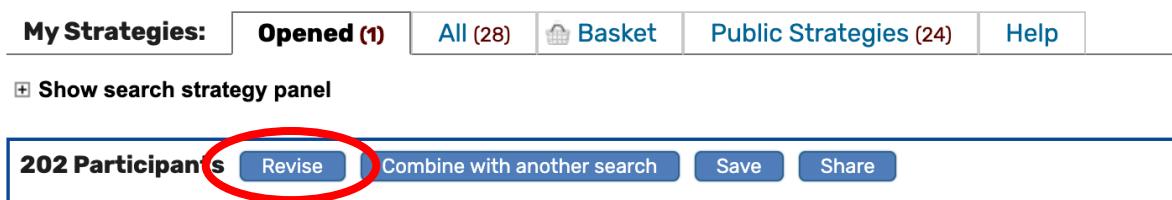
Revising your search

Exercise 2: Revising your Search Strategy

From your previous search, you now know that there were 202 participants in the MAL-ED study that had at least 1 episode of diarrhea during follow-up that was at least 5 days in duration and was preceded by a positive lab result of *Cryptosporidium* that occurred within the previous 2 weeks. How many participants may have experienced the same 5+ day diarrheal episode, except that they tested positive for *Campylobacter* instead of *Cryptosporidium* in the previous 2 weeks? To do this you can either rerun the entire search and change the microorganism of interest, or, you can simply revise this previous search. We will now work together to revise the previous search.

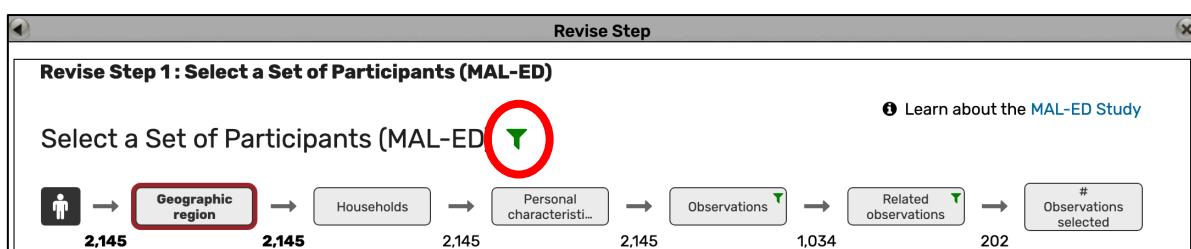
Question 2: Revise your previous search. How many participants in the MAL-ED study tested positive for *Campylobacter* within the 14 days before a diarrheal episode that was five or more days in duration?

1. From the Results Table, click on the blue “Revise” button on the top left side. If you are still on the Search Wizard page, first click on the blue “View 202 Participants” to navigate to the Results Table.

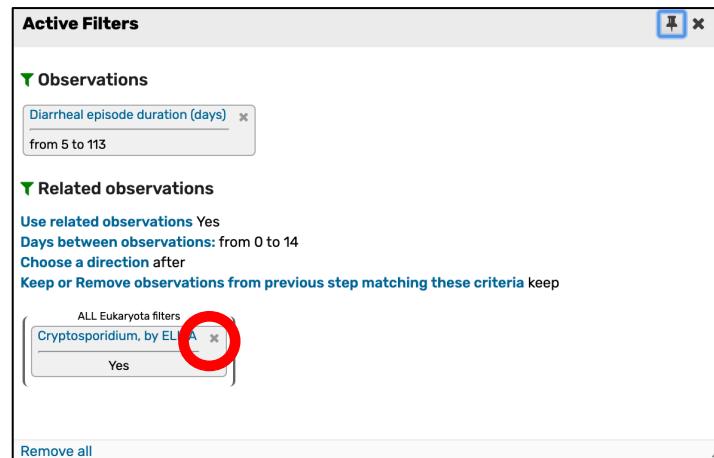


The screenshot shows a navigation bar with tabs: My Strategies (Opened 1), All (28), Basket, Public Strategies (24), and Help. Below the bar is a link to Show search strategy panel. The main area displays '202 Participants' with a 'Revise' button highlighted by a red oval. Other buttons visible are 'Combine with another search', 'Save', and 'Share'.

2. Once you click on the revise button, a popup will appear that includes the Search Wizard with the parameters you previously selected. We need to remove the “*Cryptosporidium*, by ELISA” selection criteria. Click on the green “Filter” icon next to “Select a Set of Participants (MAL-ED)” on the top left of the Search Wizard within the “Revise Step” popup. This will open up a window that lists all of the active filters, which were applied in the previous search.



3. In the “Active Filters” window, click on the “X” next to the “*Cryptosporidium*, by ELISA” filter to remove this criteria from the search strategy.



4. Now, click on the “Related Observations” box in the Search Wizard within the “Revise Step” popup and navigate to the variables for *Campylobacter* that are listed under the “Bacteria” sub-category of processed microbiology results. Check the “Yes” box next to “*Campylobacter*, by ELISA” in the multifilter selection table. Now how many participants have you selected?⁶

Bacteria	Remaining Related Observations	All Related Observations	Distribution	%
Aeromonas, by bacteriology	39,838 (100%)	39,838 (100%)	1344485 Related Observations have no data	
<input type="checkbox"/> No	39,112 (98%)	39,112 (98%)		(100%)
<input checked="" type="checkbox"/> Yes	726 (2%)	726 (2%)		(100%)
Campylobacter, by bacteriology	14,643 (100%)	14,643 (100%)	1369680 Related Observations have no data	
<input type="checkbox"/> No	13,745 (94%)	13,745 (94%)		(100%)

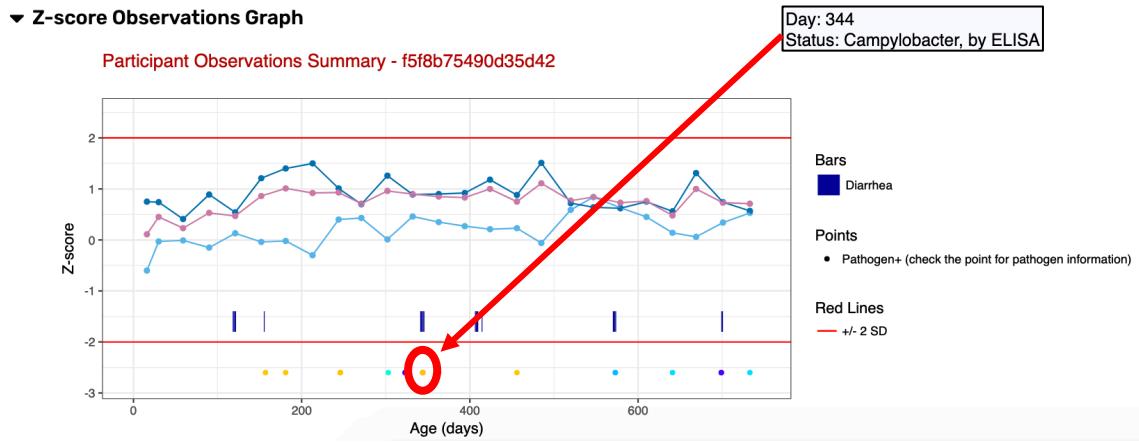
5. Note that results from two different methodologies used to detect *Campylobacter* are available. Check the “Yes” boxes next to both “*Campylobacter*, by bacteriology” and “*Campylobacter*, by ELISA” in the multifilter selection table. What happens to the number of participants that you have selected, and why has this occurred?⁷

⁶ 559 participants had a 5+ day diarrheal episode preceded by a positive *Campylobacter* ELISA result within the prior 2 weeks.

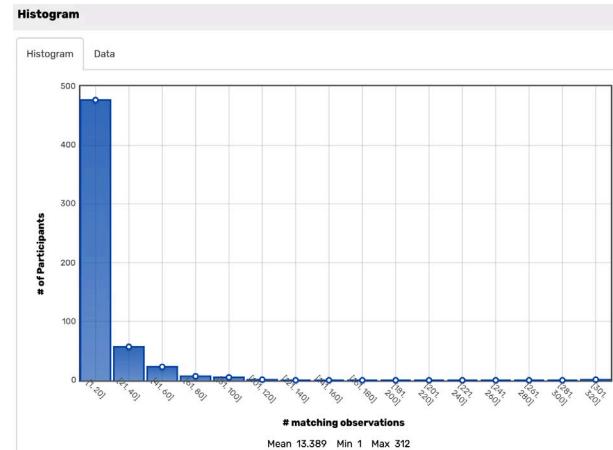
⁷ The number of selected participants has decreased from 559 to 76. This is because the participants had to test positive for *Campylobacter* by BOTH methodologies in order to be included in our selected participants

- How would you select participants who tested positive for *Campylobacter* by either methodology (bacteriology or ELISA)? How many participants meet this new search criteria?⁸
- Click on the blue “View 571 Participants” button in the Search Wizard within the “Revise Step” popup to view these results. Click through the columns and on the z-score curves to explore this dataset. Choose any participant and confirm where in their follow-up period these observations occurred.

Participant: f5f8b75490d35d42



- Look at the histogram for “# matching observations.” Compared our previous search looking at diarrhea in participants with *Cryptosporidium* infections, it appears that not only were there more Participants who had a diarrheal episode of at least 5 days following a *Campylobacter* infection but these events also tended to occur more often.



⁸ To select participants who test positive for *Campylobacter* by EITHER bacteriology or ELISA, make sure to choose “any” in the statement at the top of the multifilter selection table. 571 participants are now selected.

Bacteria

Find Related Observations with **any** of the options selected below. Make sure “any” is selected from the drop-down menu

Bacteria	Remaining Related Observations	All Related Observations	Distribution	%
Aeromonas, by bacteriology	39,838 (100%)	39,838 (100%)	1344485 Related Observations have no data	
<input type="checkbox"/> No	39,112 (98%)	39,112 (98%)	(100%)	
<input type="checkbox"/> Yes	726 (2%)	726 (2%)	(100%)	
Atypical EPEC, by PCR	38,101 (100%)	38,101 (100%)	1346222 Related Observations have no data	
<input type="checkbox"/> No	35,976 (94%)	35,976 (94%)	(100%)	
<input type="checkbox"/> Yes	2,129 (6%)	2,129 (6%)	(100%)	

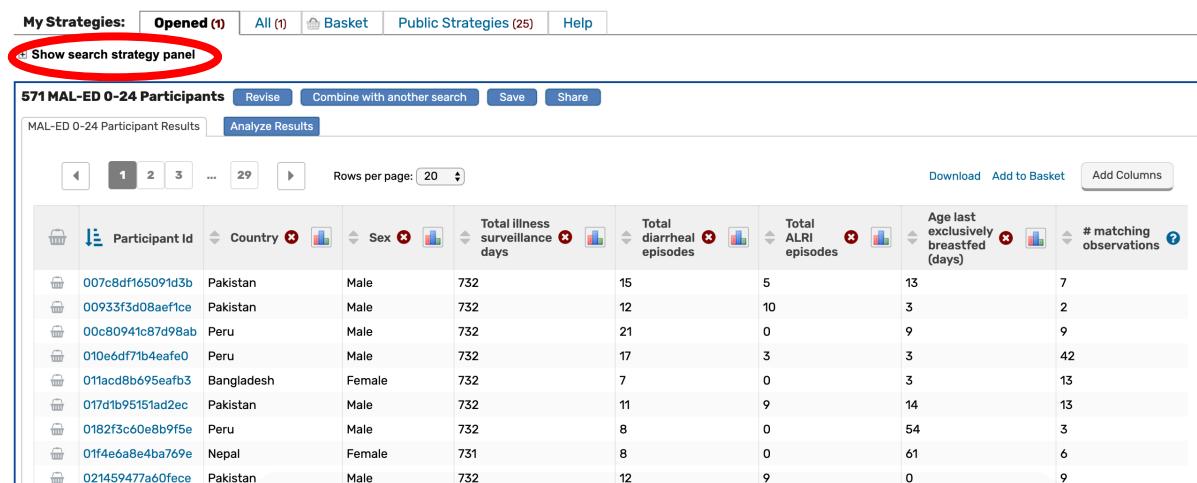
Multi-step strategies and combining searches

Exercise 3: Multi-step strategies and combining searches

In the last search, we identified the 571 MAL-ED who had 1 or more episodes of diarrhea during follow-up that were at least 5 days in duration and was preceded by a detection of *Campylobacter* in the last 2 weeks. How many of these participants also had a reduced length-for-age z-score at 2 years of age? To answer this question, you will need to employ a multi-step strategy. We will combine the results of our previous search, which identified all MAL-ED participants who had a *Campylobacter* infection followed by an episode of diarrhea lasting at least 5 days (search 1), with a new search to identify all MAL-ED participants who had a length-for-age z-score of less than -2 at 2 years of age (search 2). We will now work together to combine this new search with the search you used in Exercise 2.

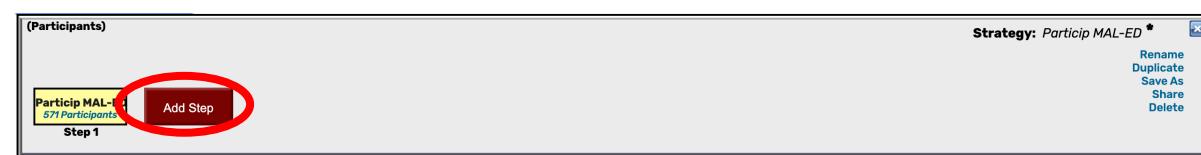
Question 3: How many participants in the MAL-ED study tested positive for *Campylobacter* within the 14 days before a diarrheal episode that was five or more days in duration and also had a length-for-age z-score of less than -2 at 2 years of age?

1. From the Results Table of the 571 MAL-ED participants who were *Campylobacter*-positive within 14 days before a diarrheal episode lasting at least 5 days, click on “Show search strategy panel” above the table.

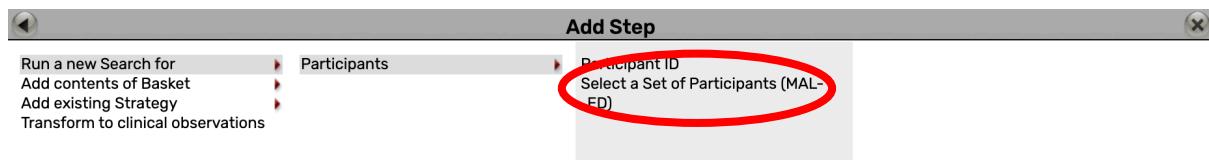


571 MAL-ED 0-24 Participants								
MAL-ED 0-24 Participant Results								
	Participant Id	Country	Sex	Total illness surveillance days	Total diarrheal episodes	Total ALRI episodes	Age last exclusively breastfed (days)	# matching observations
007c8df165091d3b	Pakistan	Male	732	15	5	13	7	
00933f3d08baef1ce	Pakistan	Male	732	12	10	3	2	
00c80941c87d98ab	Peru	Male	732	21	0	9	9	
010e6df71b4afe0	Peru	Male	732	17	3	3	42	
01acdbb695eafb3	Bangladesh	Female	732	7	0	3	13	
017db95151ad2ec	Pakistan	Male	732	11	9	14	13	
0182f3c60e8b9f5e	Peru	Male	732	8	0	54	3	
01f4e6a8e4ba769e	Nepal	Female	731	8	0	61	6	
021459477a60fce	Pakistan	Male	732	12	9	0	9	

2. In the search strategy panel, click the “Add Step” box next to the search from Question 2.



3. In the “Add Step” popup window, choose “Select a set of clinical observations (MAL-ED).”



4. The “Add Step” popup window will now open a new Search Wizard. Use this Search Wizard to obtain a list of all MAL-ED participants who had a length-for-age z-score lower than -2 at 2 years of age (search 2). Notice that there is a new section above the Search Wizard in the “Add Step” popup. This section allows you to choose how to combine this new search (search 2) with the original search (search 1). Since we want to identify all participants who had a length-for-age z-score less than -2 at 2 years of age (search 2) who **ALSO** were *Campylobacter*-positive within 14 days before a diarrheal episode lasting at least 5 days (search 1), we want to find the intersect of the two searches. Go ahead and select “1 Intersect 2.”

Add Step 2 : Select a Set of Participants (MAL-ED)

Your Step 1 found 571 Participants. To combine it with a new search, choose a combine operation, then specify the new search. [Learn about the MAL-ED Cohort](#)

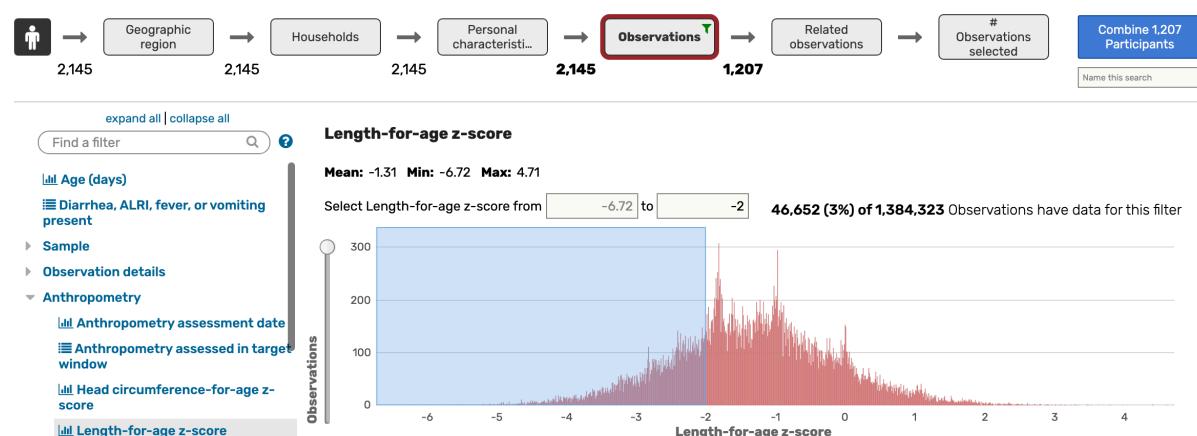
1 Intersect 2 1 Minus 2
 Union 2 2 Minus 1

Select a Set of Participants (MAL-ED)

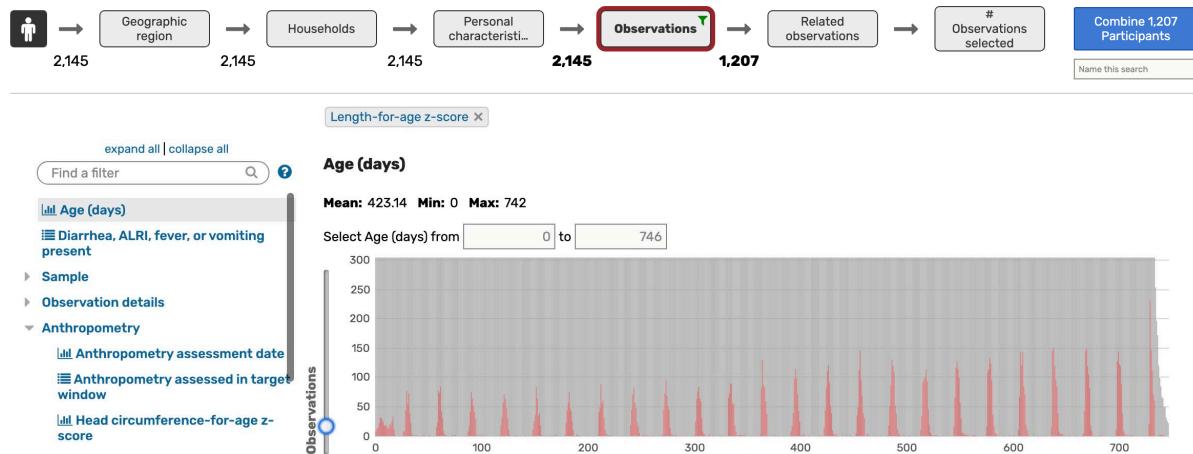
```

    graph LR
      Start(( )) --> Geographic[Geographic region]
      Geographic --> Households[Households]
      Households --> Personal[Personal characteristi...]
      Personal --> Observations[Observations]
      Observations --> Related[Related observations]
      Related --> Selected["# Observations selected"]
      Selected --> Combine[Combine 2,145 Participants]
      
```

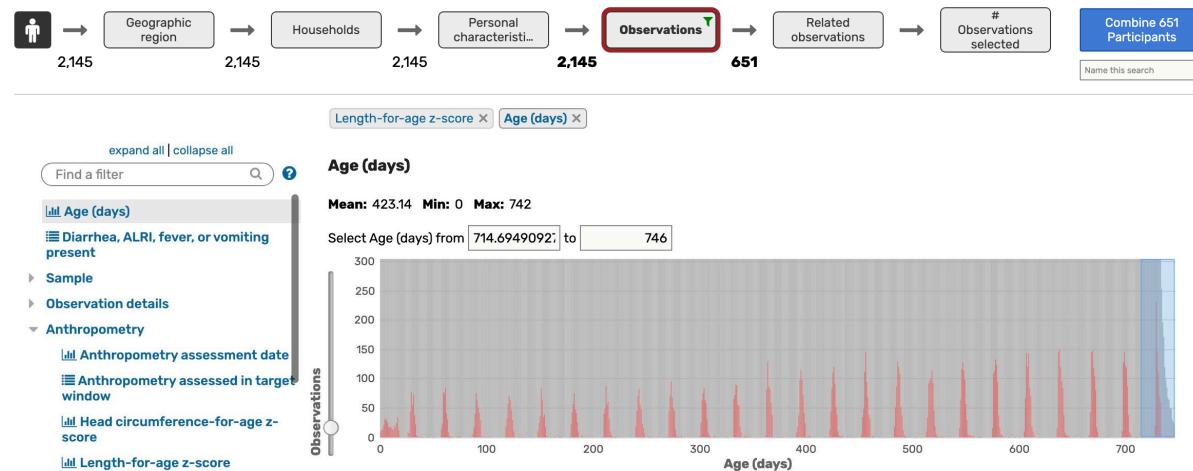
5. Now use the Search Wizard in the “Add Step” popup to perform search 2. Here, we want to identify the list of participants who had a length-for-age z-score less than -2 at 2 years of age. The first step in this search is to select participants who had a length-for-age z-score less than -2 at any point during the study. Z-scores are sorted under the “Observations” step in the Search Wizard. Remember that you can use “Find a filter” to quickly search for variables of interest.



6. We still need to limit the selected length-for-age z-scores to those that were collected when the participant was 2 years of age. Navigate to the “Age (days)” variable in the “Observations” step of the Search Wizard. Use the scroll bar on the left side of the graph to zoom in on the data. For this variable, notice that there are a series of red peaks on a grey background. Why do you think the data looks like this?⁹



7. Select data that was collected at approximately 2 years of age. There is variation on exactly when the 2 year timepoint of anthropometry data was collected, so be sure to account for this variation when choosing “Age (days).”



8. Search 2 determined that there were 651 MAL-ED participants who had a length-for-age z-score less than -2 at 2 years of age. Click on the blue “Combine 651 Participants” box in the Search Wizard to combine the results of search 2 with search 1.

The dialog box is titled 'Add Step 2 : Select a Set of Participants (MAL-ED)'. It says 'Your Step 1 found 571 Participants. To combine it with a new search, choose a combine operation, then specify the new search.' It includes a 'Learn about the MAL-ED Cohort' link. Below are four combine operation options: '1 Intersect 2', '1 Minus 2', '1 Union 2', and '2 Minus 1'. The 'Combine 651 Participants' button is highlighted with a red oval.

⁹ From the information on the MAL-ED Data Set page, we can determine that anthropometry variables such as length-for-age z-scores were measured monthly. The peaks represent every month of anthropometry data collection, with some variation around the exact day of age when the data was collected.

9. How many participants in the MAL-ED study tested positive for *Campylobacter* within the 14 days before a diarrheal episode that was five or more days in duration and also had a length-for-age z-score of less than -2 at 2 years of age?¹⁰
-

Independent work

Exercise 4: Use data from various studies on ClinEpiDB.org to answer the following questions – Independent work (45 minutes)

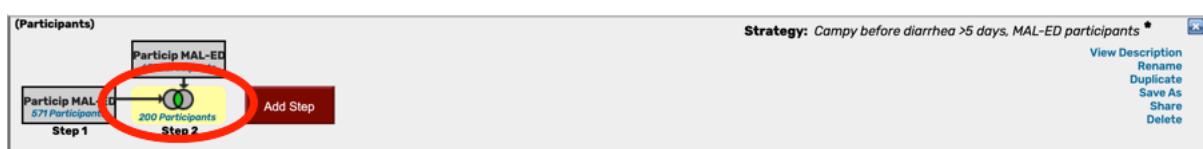
In this exercise, you will have an opportunity to explore the following questions on your own. If you have questions or get stuck, raise your hand and a moderator will come over to help you. Remember to clear any applied selection criteria before approaching each question. Try to start at the left side 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 Participant-level, Observation-level, Household-level, and Light Trap-level searches on the data, and think about whether the question is asking you before initiating a search.

Question 4: What is the average age of first asymptomatic parasitemia in the PRISM cohort?

Whether someone infected with *Plasmodium falciparum* develops malaria is based on host genetics, parasite strain, and previous parasite exposure, which can eventually lead to immune system protection. Age at first asymptomatic parasitemia can provide a clue about the transmission intensity at a particular site: people who live in areas with higher *P. falciparum* transmission rates experience asymptomatic parasitemia at earlier ages.

If we want to determine what age participants first experienced asymptomatic parasitemia, what type of search do we need to do (Participant, Household, Observation, or Light trap)?¹¹ To answer this question, you will initiate an Observation-level search and employ the “Related observations” filter. You will use histograms from the results page to find the average age of first asymptomatic parasitemia, and you will revise your search based on other key filters of interest. If you have extra time, you can discover how to use multi-step strategies to refine your search even further.

¹⁰ There were 200 participants that were returned in both search 1 and search 2.



¹¹An Observation-level search will allow us to determine the age at a specific observation where a child had asymptomatic parasitemia. Using the special “Related observation” step in the Search Wizard will allow us to remove observations that are not the first observation of asymptomatic parasitemia.

1. Navigate to <http://clinepidb.org> in your browser. Start an Observation-level search of the PRISM study by clicking the “Observation” icon on either the study card or the drop-down menu under “Search a Study.”

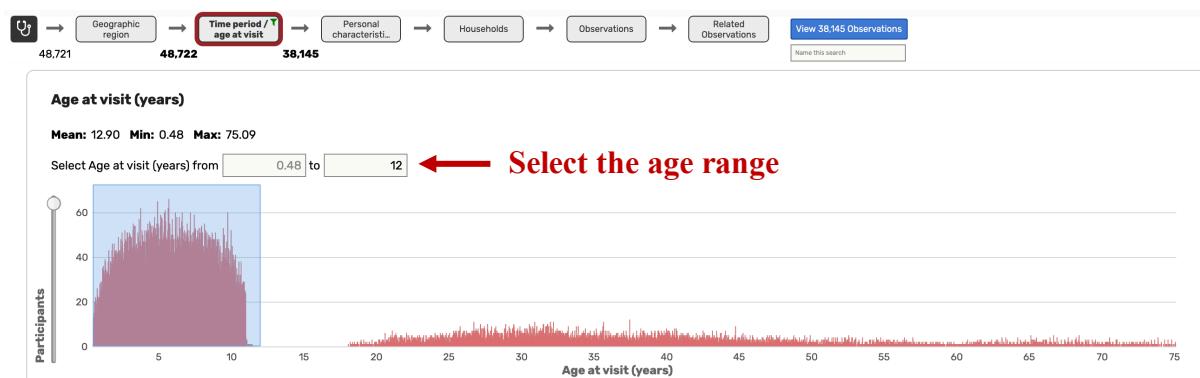
Select a Set of Clinical Observations (PRISM)

Learn about the PRISM Cohort

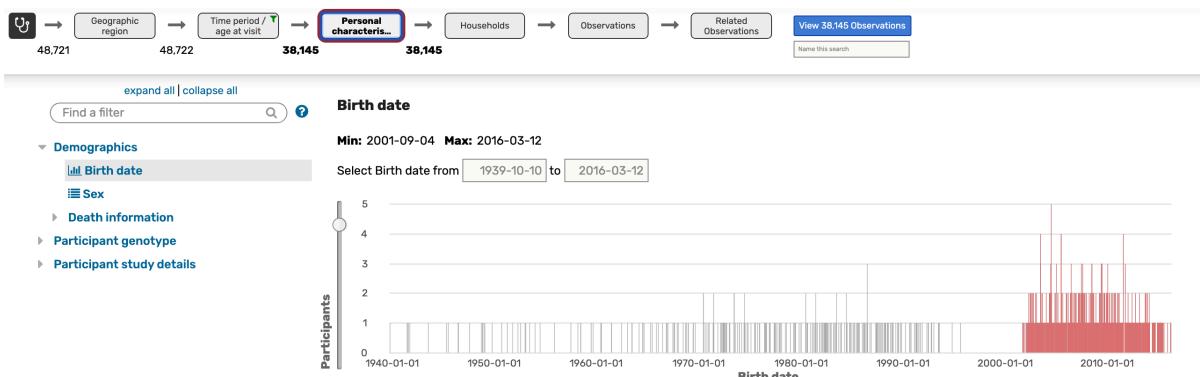
View 48,722 Observations

Name this search

2. We want to filter the data for children. Where can you find variables related to age?¹²
3. Click on “Time period/age at visit” in the Search Wizard and select children under 12. Remember, you can select data by either entering the range into the indicated boxes or clicking and dragging your mouse over the range you want to select. How many observations remain?¹³



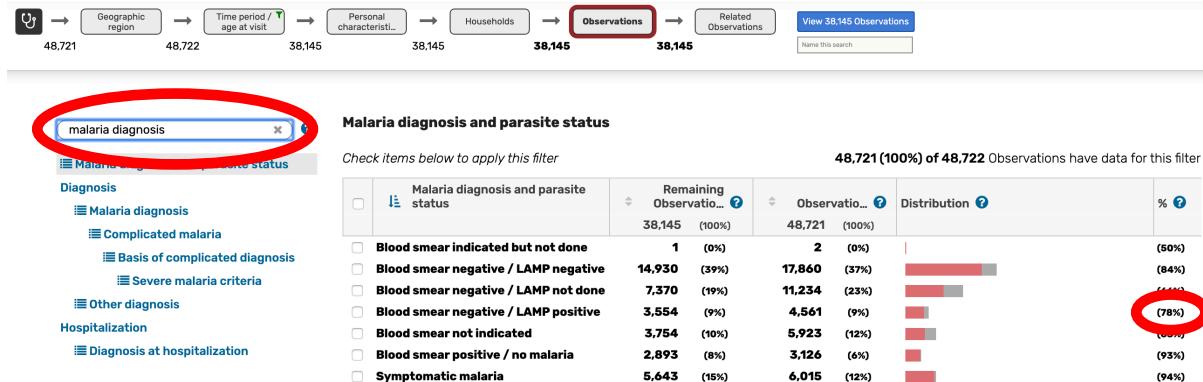
4. Move to the “Personal characteristics” step in the Search Wizard. It will open to “Birth date” by default. Note that all birth dates before 2000 are in grey, while all birth dates after 2000 are in pink. This is because we filtered the dataset to select participants younger than 12 years old. Only participants born in 2001 or later are selected by our previously-applied criteria. As we go through the Search Wizard, remember that all data that meet previous selection criteria will be shown in pink, while data that do not meet the criteria will be shown in grey.



¹² Age-related variables actually appear in 3 places in the PRISM cohort study, under “Time period/age at visit”, “Personal characteristics”, and “Observations.” In this case we will use the variable under “Time period/age at visit.”

¹³ 38,145 Observations remain after limiting our search to observations occurring before 12 years of age.

5. Now let's find malaria variables. Click on "Observation" in the Search Wizard and look through the available variables. For this example, we want to use "Malaria diagnosis and parasite status". You can find a specific variable by typing its name into the "Find a filter" search box.

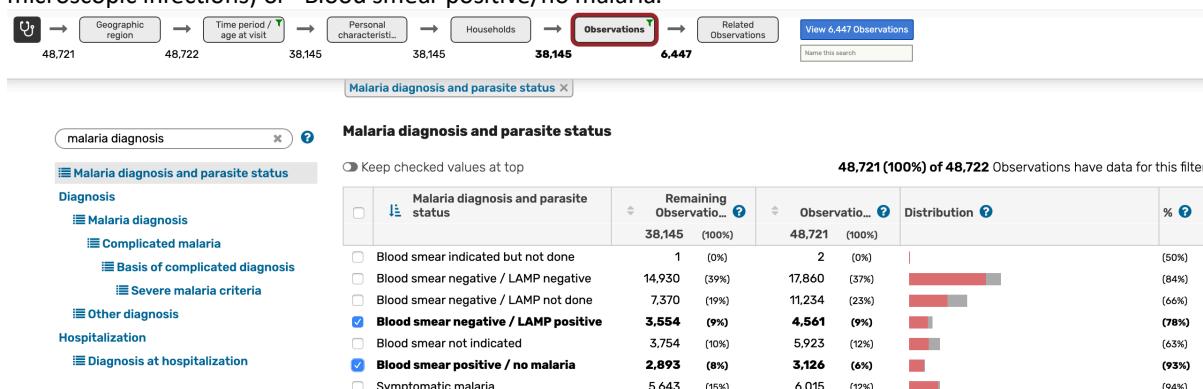


6. Similar to the Date of birth histogram above, the Distribution column graphs data that meet filter criteria are in pink and data that don't meet the filter criteria are in grey. Can you interpret what the 78% circled in red above means? Remember, you can mouse over the question mark at the top of the column to learn more.¹⁴
7. How many observations of "Blood smear positive / no malaria" were made in participants 12 years of age or younger? How many observations of "Blood smear positive / no malaria" were made in participants older than 12 years of age?¹⁵
8. We are interested in identifying all observations where a child was infected with *Plasmodium falciparum* but did not show symptoms. Which values for "Malaria diagnosis and parasite status" should we select?¹⁶

¹⁴ That percentage indicates the percent of Remaining Observations out of all Observations. It shows what percent of the bar in the Distribution column is pink, or has been selected via the filters. In this particular case, it tells you that 78% of all observations that were "Blood smear negative / LAMP positive" occurred in children.

¹⁵ 2,893 observations of "Blood smear positive / no malaria" were made in participants 12 years of age or younger. 233 (3,126 – 2,893) observations of "Blood smear positive / no malaria" were made in participants older than 12 years of age.

¹⁶We are interested in children who were either "Blood smear negative/LAMP positive" (which indicates sub-microscopic infections) or "Blood smear positive/no malaria."



9. Now let's move to the "Related Observations" step in the Search Wizard. Once you go to that step, everything is greyed out until you check the box to enable the advanced related observations filter. Click the check box to enable the filter, which makes all of the variables from the "Observations" step available. The key difference between the "Observations" and the "Related Observations" steps is that the "Related Observations" filter works by allowing you to keep or remove observations you already selected for based on a second set of observations that you define in this step.

Remove Observations that are to days before the Related Observations specified below

Enable the advanced Related Observations filter below. It allows you to restrict Observations by relating them to your choice of Related Observations.

Malaria diagnosis and parasite status

Check items below to apply this filter

48,721 (100%) of 48,722 Observations have data for this filter

	Remaining Observatio... <small>?</small>	Observatio... <small>?</small>	Distribution <small>?</small>	% <small>?</small>
48,721 (100%)	48,721 (100%)			
Blood smear indicated but not done	2 (0%)	2 (0%)		(100%)
Blood smear negative / LAMP negative	17,860 (37%)	17,860 (37%)		(100%)
Blood smear negative / LAMP not done	11,234 (23%)	11,234 (23%)		(100%)
Blood smear negative / LAMP positive	4,561 (9%)	4,561 (9%)		(100%)
Blood smear not indicated	5,923 (12%)	5,923 (12%)		(100%)
Blood smear positive / no malaria	3,126 (6%)	3,126 (6%)		(100%)
Symptomatic malaria	6,015 (12%)	6,015 (12%)		(100%)

10. To answer this question, we are trying to identify the **FIRST** observation of asymptomatic parasitemia in children. Right now we have selected **ALL** observations where asymptomatic parasitemia was detected in children. We need to remove all observations of asymptomatic parasitemia except for the first one.

To start, let's define what "Related Observation" we're interested in. In this case, we want to differentiate between the first asymptomatic infection and all other observations of asymptomatic infection, so we will define our "Related Observation" the same way we defined our "Observation." Click on the "Malaria diagnosis and parasite status" variable in the "Related Observations" step in the Search Wizard and select the values of "Blood smear negative/LAMP positive" and "Blood smear positive/no malaria."

Remove Observations that are to days before the Related Observations specified below

Enable the advanced Related Observations filter below. It allows you to restrict Observations by relating them to your choice of Related Observations.

Malaria diagnosis and parasite status

Keep checked values at top

48,721 (100%) of 48,722 Observations have data for this filter

	Remaining Observatio... <small>?</small>	Observatio... <small>?</small>	Distribution <small>?</small>	% <small>?</small>
48,721 (100%)	48,721 (100%)			
Blood smear indicated but not done	2 (0%)	2 (0%)		(100%)
Blood smear negative / LAMP negative	17,860 (37%)	17,860 (37%)		(100%)
Blood smear negative / LAMP not done	11,234 (23%)	11,234 (23%)		(100%)
Blood smear negative / LAMP positive	4,561 (9%)	4,561 (9%)		(100%)
Blood smear positive / no malaria	3,126 (6%)	3,126 (6%)		(100%)
Symptomatic malaria	6,015 (12%)	6,015 (12%)		(100%)

11. Now we specify how to treat other observations in relation to this observation. Since we want to retain only the first observation of asymptomatic parasitemia for each child, we can choose to **Remove** observations **1** to **9999** days **after** the related observation we specified below.

Enable the advanced Related Observations filter below. It allows you to restrict Observations by relating them to your choice of Related Observations.

Remove Observations that are **1** to **9999** days **after** the Related Observations specified below

Malaria diagnosis and parasite status

	Remaining Observatio...	Observatio...	Distribution	%
<input type="checkbox"/> Malaria diagnosis and parasite status	48,721 (100%)	48,721 (100%)		
<input type="checkbox"/> Blood smear indicated but not done	2 (0%)	2 (0%)		(100%)
<input type="checkbox"/> Blood smear negative / LAMP negative	17,860 (37%)	17,860 (37%)		(100%)
<input type="checkbox"/> Blood smear negative / LAMP not done	11,234 (23%)	11,234 (23%)		(100%)
<input checked="" type="checkbox"/> Blood smear negative / LAMP positive	4,561 (9%)	4,561 (9%)		(100%)
<input type="checkbox"/> Blood smear not indicated	5,923 (12%)	5,923 (12%)		(100%)
<input checked="" type="checkbox"/> Blood smear positive / no malaria	3,126 (6%)	3,126 (6%)		(100%)
<input type="checkbox"/> Symptomatic malaria	6,015 (12%)	6,015 (12%)		(100%)

12. 854 observations of the first asymptomatic parasitemia in children remain. Click on the blue “View 854 Participants” button in the Search Wizard to visit the results page.

13. Remember that you can add and remove columns, sort columns, and view histograms of the selected data on the results page. Make sure that your Results Table has a column for

Sort **Remove column** **View histogram** **Add columns**

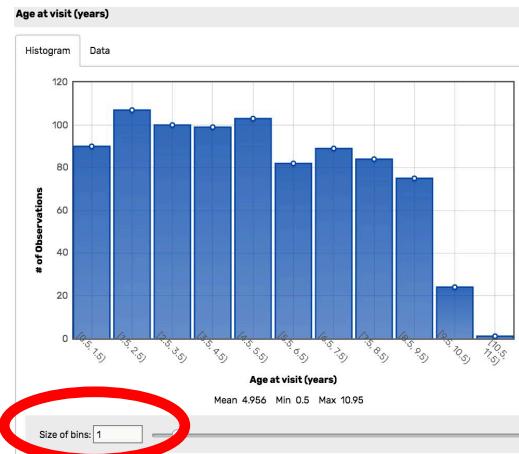
Age at visit (years)

14. To determine what the **AVERAGE** age at first asymptomatic parasitemia is, open the histogram for “Age at visit (years).” What is the average age?¹⁷



¹⁷ The average age at first asymptomatic infection is 4.956.

15. Change the size of the bins on the histogram to 1. How many children had their first observed asymptomatic infection at 1.5-2.5 years old?¹⁸ Hint: hover over the bars to see additional information.



Question 5: We know that the three different PRISM sites were selected to represent different transmission settings. How would you expect the average age at first asymptomatic visit in Nagongera, a high transmission setting, to compare to Walukuba, a lower transmission setting?¹⁹ Can you prove your hypothesis using the data on ClinEpiDB.org?

Remember that instead of starting a new search from scratch, we can edit a search strategy by clicking the “Revise” button from the results page. To answer this question, start by first duplicating the search and then revising it, so we can come back to this search easily later.

1. Click on the “+” sign next to “Show search strategy panel” in the top left of the results page. This will open a panel that shows your current search at the top and any other searches you have open below it.

2. We will first name the search strategy used to answer Question 4 so that we can return to it later. You can rename the search by clicking on its name or by clicking “Rename.” Name this search “First asymptomatic.”

¹⁸ 107 children had their first observed asymptomatic parasitemia at 1.5 to 2.5 years of age.

¹⁹ Nagongera is expected to have a lower average age at first asymptomatic parasitemia since children likely develop immunity faster due to more frequent infection.

3. Click on “Duplicate” on the right-hand side of the search strategy panel. You will see a pop-up asking if you want to make a copy of the strategy. Click “OK”.

4. Now you should see two searches in the search strategy panel. Rename the new strategy “Nagongera.”

5. Now click the blue “Revise” button. Note that the results table shows data from whichever box is highlighted yellow in the search strategy panel. The Nagongera search is currently highlighted yellow, so clicking “Revise” will allow you to edit this search.

Observation Id	Visit type	Visit date	Age at visit (years)	Malaria diagnosis and parasite status	# relative observations	Min days between observations
100320227	Scheduled visit	2015-05-24 00:00:00.0	8.26	Blood smear positive / no malaria	N/A	N/A
100619044	Scheduled visit	2012-02-25 00:00:00.0	8.68	Blood smear negative / LAMP positive	N/A	N/A

6. Within the “Revise Step” popup, select “Nagongera” for the “Sub-county in Uganda” variable in the “Geographic Region” step in the Search Wizard. You can click “Refresh counts” to see how this new filter impacts the other filters you already put in place. How many observations have you selected now?²⁰

²⁰ 349 observations of first asymptomatic infections are from children at the Nagongera site.

Revise Step

Revise Step 1: Select a Set of Clinical Observations (PRISM)

Learn about the PRISM Cohort

Select a Set of Clinical Observations (PRISM)

Geographic region

48,721 → 23,044 → Time period / age at visit → Personal characteristics → Households → Observations → Related Observations → View Observations

Your Geographic region filters reduce 48,721 Observations to 23,044

Sub-county in Uganda

Keep checked values at top

	Remaining Households	Households	Distribution	%
<input type="checkbox"/> Sub-county in Uganda	331 (100%)	331 (100%)		
<input type="checkbox"/> Kihibi	107 (32%)	107 (32%)		(100%)
<input checked="" type="checkbox"/> Nagongera	107 (32%)	107 (32%)		(100%)
<input type="checkbox"/> Walukuba	117 (35%)	117 (35%)		(100%)

7. Now click the blue “View Observations” button to return to the results page. What is the average age at the visit where asymptomatic parasitemia was first detected now?²¹
8. We would also like to see what the average age at first asymptomatic parasitemia is for children in Walukuba. Revise your search strategy. What is the average age at the visit where asymptomatic parasitemia was first detected for Walukuba?²²
9. Do you think this is an accurate estimate of first age of asymptomatic infection, why or why not? One thing to consider is that some of the children were enrolled when they were older. How might this impact the data?²³ How could we adapt the search to correct for this?²⁴

²¹ Now the average age at visit is 4.533 years, which is lower than the original value for data from all 3 sites. This is what we would expect since Nagongera has a higher transmission rate, so children are exposed to *Plasmodium* more frequently and build immunity faster.

²² 4.956 years in Walukuba.

²³ Older children may have already experienced their first asymptomatic infection. In this case the results are telling us when they experienced their first asymptomatic infection once they were enrolled in the study, but not their actual first asymptomatic infection.

²⁴ To correct for this, we can limit our participants to younger children. There is a filter for “Age at enrollment” under the step “Personal characteristics”. Try limiting this and see what you find.

Question 6: Consider the possibility that some observations of asymptomatic parasitemia may have been detected early on in the course of infection, before becoming symptomatic malaria. Out of the 854 observations of first asymptomatic infections in children enrolled in the PRISM cohort, how many did not develop into symptomatic malaria in the 2 weeks following detection?

To answer this question, we will want to adapt our search from Question 4 (named “First asymptomatic”) to remove observations where the participant converted to symptomatic infection shortly after detection of asymptomatic parasitemia (within 2 weeks). To remove infections that became symptomatic, we will need to use a multi-step strategy. This requires adding a separate step where we identify all asymptomatic infection observations that became symptomatic in 2 weeks. Then, we will subtract those results from the results of our original search for first asymptomatic infections.

10. Start by clicking the “Add Step” box for the search from Question 4. We named this search “First asymptomatic”, and it returned 854 observations.

The screenshot shows the Clinical Observations interface with two search strategies. The top section displays the 'Nagongera' strategy with 349 observations. The bottom section displays the 'First asymptomatic' strategy with 854 observations. A red circle highlights the 'Add Step' button in the bottom section.

11. In the “Add Step” popup window, choose “Select a set of clinical observations (PRISM).”

The screenshot shows the 'Add Step' popup window. The 'Observations' dropdown is selected, and the option 'Select a Set of Clinical Observations (PRISM)' is highlighted with a red circle.

12. Notice that the panel that opens up looks exactly the same as the Search Wizard you used earlier, except for the new section at the top. This section allows you to choose how to combine this search (search 2) with the original search (search 1). To answer Question 6, we want to identify all observations that became symptomatic (search 2) and remove them from the original search (search 1) . Which combine operation do you think we should use?²⁵ Go ahead and check the box next to that combine operation now.

The screenshot shows the 'Add Step 2 : Select a Set of Clinical Observations (PRISM)' panel. It displays a message: 'Your Step 1 found 854 Participants. To combine it with a new search, choose a combine operation, then specify the new search.' Below this, there are four radio button options for combining search 1 and search 2:

- 1 Intersect 2
- 1 Minus 2
- 1 Union 2
- 2 Minus 1

²⁵ 1 Minus 2

13. The first part of search 2 should be similar to what we did previously in search 1. We want to identify children with asymptomatic infection.

- Under the “Time period/age at visit” step in the Search Wizard, select participants 12 years of age or younger.
- Then under the “Observations” step, click on “Malaria diagnosis and parasite status” and select the values associated with asymptomatic infection (“Blood smear negative/LAMP positive” and “Blood smear positive/no malaria”).

If you click on the green filter icon above the Search Wizard, a pop-up will appear that shows the filters you applied. It should look like this:

Select a Set of Clinical Observations (PRISM) ▾

Geographic region → Time period / age at visit → Personal characteristi... → Households → **Observations** → Related Observations → Combine 6,447 Observations

48,721 48,722 38,145 38,145 38,145 6,447 Name this search

Active Filters

Time period / age at visit

Age at visit (years) less than 12

Observations

Malaria diagnosis and parasite status

Blood smear positive / no malaria, Blood smear negative / LAMP positive

14. Now go to “Related observations.” We want to keep all observations that become symptomatic for step 2.

- Which value should you select for “Malaria diagnosis and parasite status”?²⁶
- We want to keep observations that became symptomatic up to 14 days after our original observation. How can we specify this under “Related observations”?²⁷

Select a Set of Clinical Observations (PRISM) ▾

Geographic region → Time period / age at visit → Personal characteristi... → Households → Observations → **Related Observations** → Combine 634 Observations

48,721 48,722 38,145 38,145 38,145 6,447 Name this search

Enable the advanced Related Observations filter below. It allows you to restrict Observations by relating them to your choice of Related Observations.

Keep 1 to 14 days before the Related Observations specified below

Malaria diagnosis and parasite status

Malaria diagnosis and parasite status

Keep checked values at top 48,721 (100%) of 48,722 Observations have data for this filter

	Malaria diagnosis and parasite status	Remaining Observatio...	Observatio...	Distribution	%
	48,721 (100%)	48,721 (100%)	48,721 (100%)		
<input type="checkbox"/>	Blood smear indicated but not done	2 (0%)	2 (0%)		(100%)
<input type="checkbox"/>	Blood smear negative / LAMP negative	17,860 (37%)	17,860 (37%)		(100%)
<input type="checkbox"/>	Blood smear negative / LAMP not done	11,234 (23%)	11,234 (23%)		(100%)
<input type="checkbox"/>	Blood smear negative / LAMP positive	4,561 (9%)	4,561 (9%)		(100%)
<input type="checkbox"/>	Blood smear not indicated	5,923 (12%)	5,923 (12%)		(100%)
<input type="checkbox"/>	Blood smear positive / no malaria	3,126 (6%)	3,126 (6%)		(100%)
<input checked="" type="checkbox"/>	Symptomatic malaria	6,015 (12%)	6,015 (12%)		(100%)

²⁶ Choose “Symptomatic malaria”

²⁷ Keep observations that are 1 to 14 days before the related observations specified below

15. There should be 634 observations selected in step 2. Go to the results page by clicking the blue “Combine 634 observations” button at the end of the Search Wizard.

Select a Set of Clinical Observations (PRISM) ▾



16. How many observations of first asymptomatic infection are there after removing asymptomatic infections that converted to symptomatic infections within two weeks?²⁸

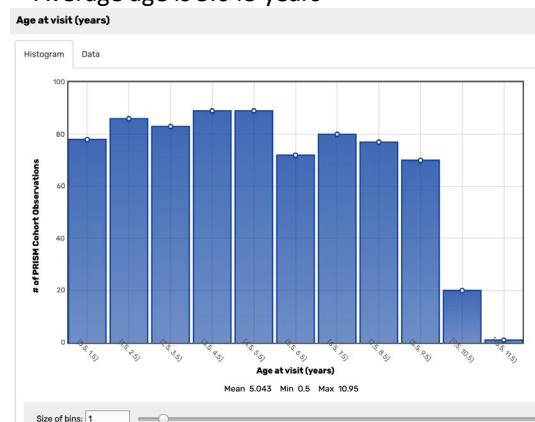
17. To view the Results Table for the 745 observations of first asymptomatic parasitemia that did not convert to symptomatic malaria, click on the blue text “745 Observations” in the strategy panel. Make sure that this area is now highlighted yellow. Also make sure that your Results Table is displaying the data you want; the title of the table should now read “745 PRISM Cohort Observations from Step 2.”

Observation Id	Visit type	Visit date	Age at visit (years)	Malaria diagnosis and parasite status
100320227	Scheduled visit	2015-05-24 00:00:00.0	8.26	Blood smear positive / no malaria
100619044	Scheduled visit	2012-02-25 00:00:00.0	8.68	Blood smear negative / LAMP positive
100719044	Scheduled visit	2012-02-25 00:00:00.0	5.21	Blood smear negative / LAMP positive
100918844	Enrollment	2011-07-29 00:00:00.0	7.14	Blood smear negative / LAMP positive
101218845	Enrollment	2011-08-09 00:00:00.0	7.15	Blood smear negative / LAMP positive

18. Look at the results table. What's the average age at visit? Remember you can add columns using the “Add column” button at the top right of the table.²⁹ How did the average age change when compared to our first search? Did you expect this?

²⁸ 745 observations

²⁹ Average age is 5.043 years



Question 7: find all participants with faltering growth after campy infection in MALED

i. <https://clinepidb.org/ce/im.do?s=b4a00d4c79303045>