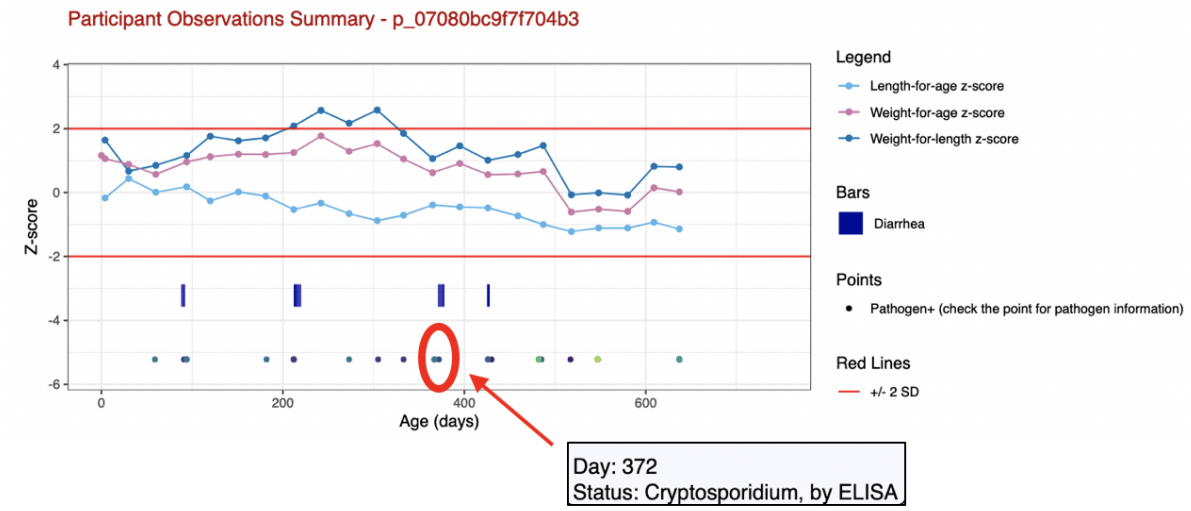
Complex Epidemiology Queries
Guided Walk-Through

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.



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.

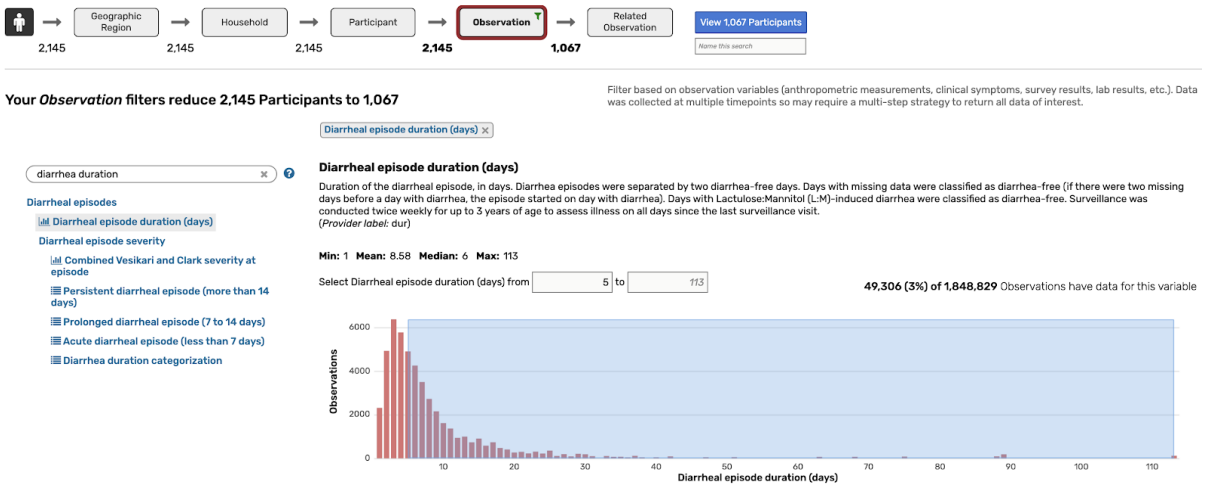
You can also see that this participant had multiple instances of positive laboratory test results for microbiology, indicated by the colored dots at the bottom of the graph (different colored dots represent the detection of different pathogens). One question you might want to ask is, **"Did participants 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.

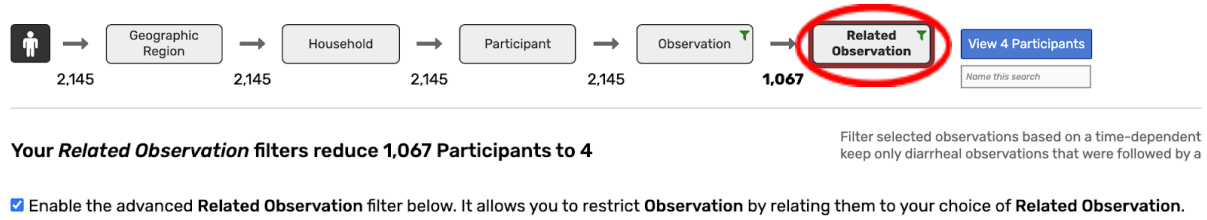
You can do this by using the Related Observations functionality! We will start with a very simple related observations query. For this exercise, to start, we will modify as few filters as possible to keep things straightforward.

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. 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."



3. 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]

4. To determine the number of 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 this question, we need to identify participants who tested positive for *Cryptosporidium* within the 2 weeks ***BEFORE*** a diarrheal episode lasting at least 5 days.

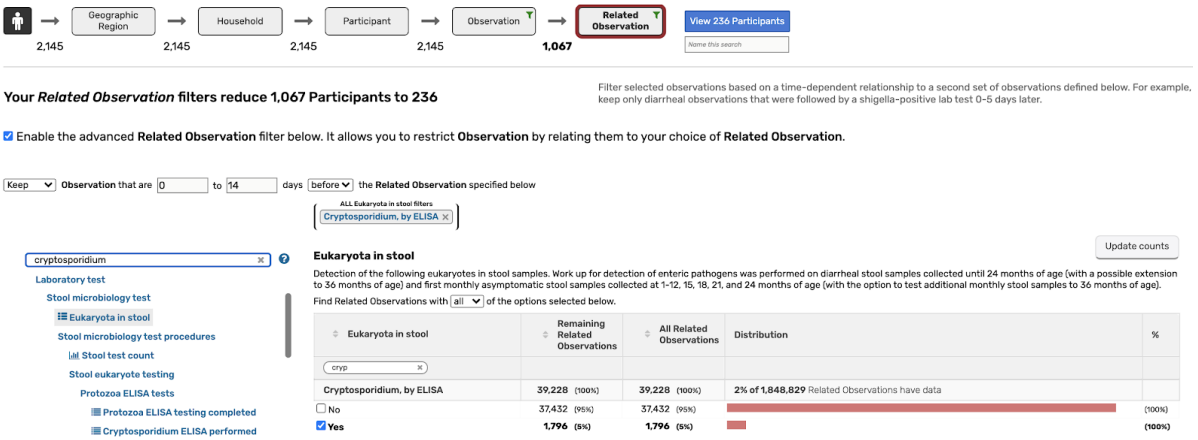
   In the Observations step in the Search Wizard, we previously selected participants who had diarrheal episodes that lasted for 5+ days. We now want to narrow down the selected participants further, and keep only the selected participants who also had a positive lab test for *Cryptosporidium.* However, 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.

   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).*

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. 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.

[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.

[2] The sentence should be completed to read: "**Keep** observations that are **0 to 14** days **before** the Related Observation specified below."



6. 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]

Multi-step strategies and combining searches

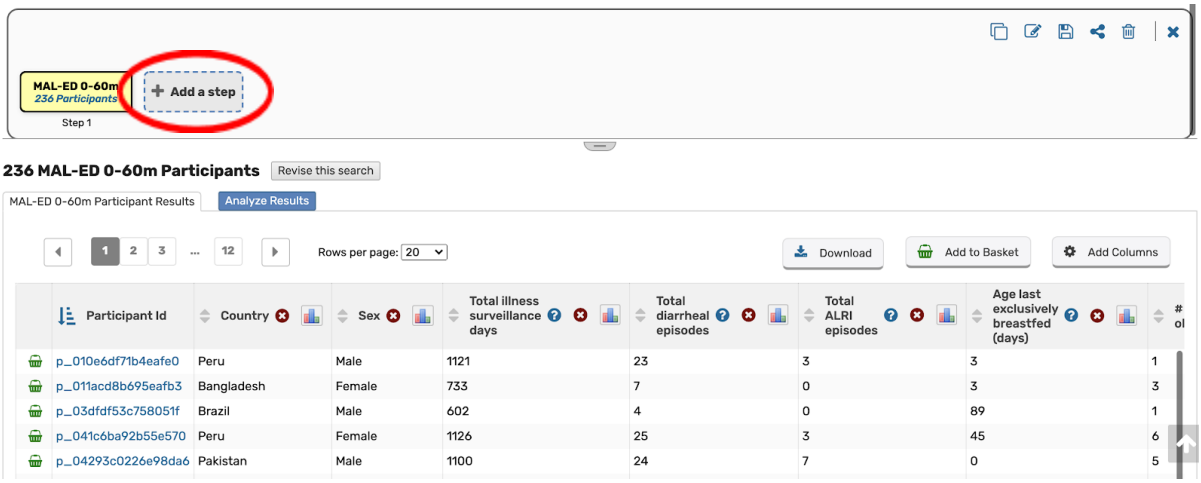
In the last search, we identified MAL-ED participants who had 1 or more episodes of diarrhea that were at least 5 days in duration and that were preceded by a detection of *Cryptosporidium* in the last 2 weeks. How many of these participants also had a reduced length-for-age z-score at 2 years of age?

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 *Cryptosporidium* 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)**.

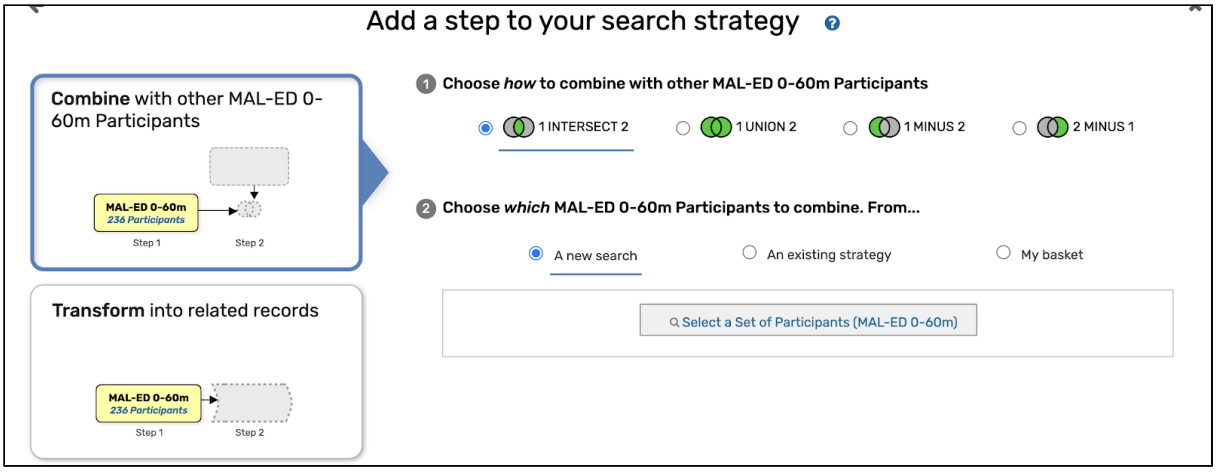
Question 2: How many participants in the MAL-ED study tested positive for *Cryptosporidium* within the 14 days before a diarrheal episode that was five or more days in duration and also had a length-for-age z-score of less than -2 at 2 years of age?

[3] Using the "Related Observations" filter, the remaining participants was reduced from 1,067 participants who had a diarrheal episode at least 5 days in length to 236 participants who had a 5+ day diarrheal episode preceded by a positive *Cryptosporidium* result within the previous 2 weeks.

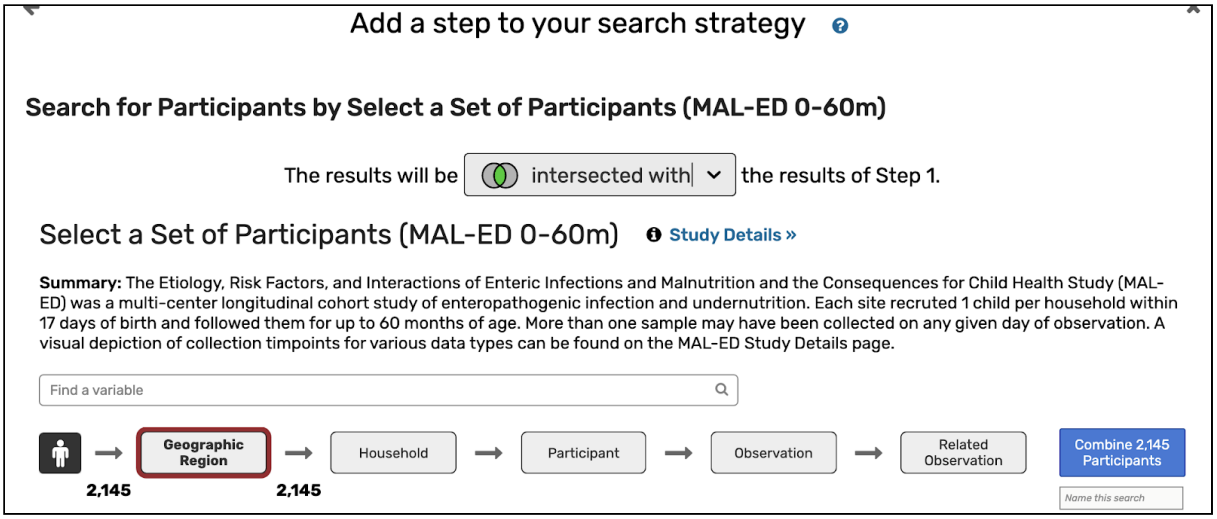
1. From the Results Table of the 236 MAL-ED participants who were *Cryptosporidium*-positive within 14 days before a diarrheal episode lasting at least 5 days, navigate to the search strategy panel above the table and click the “Add a step” button.



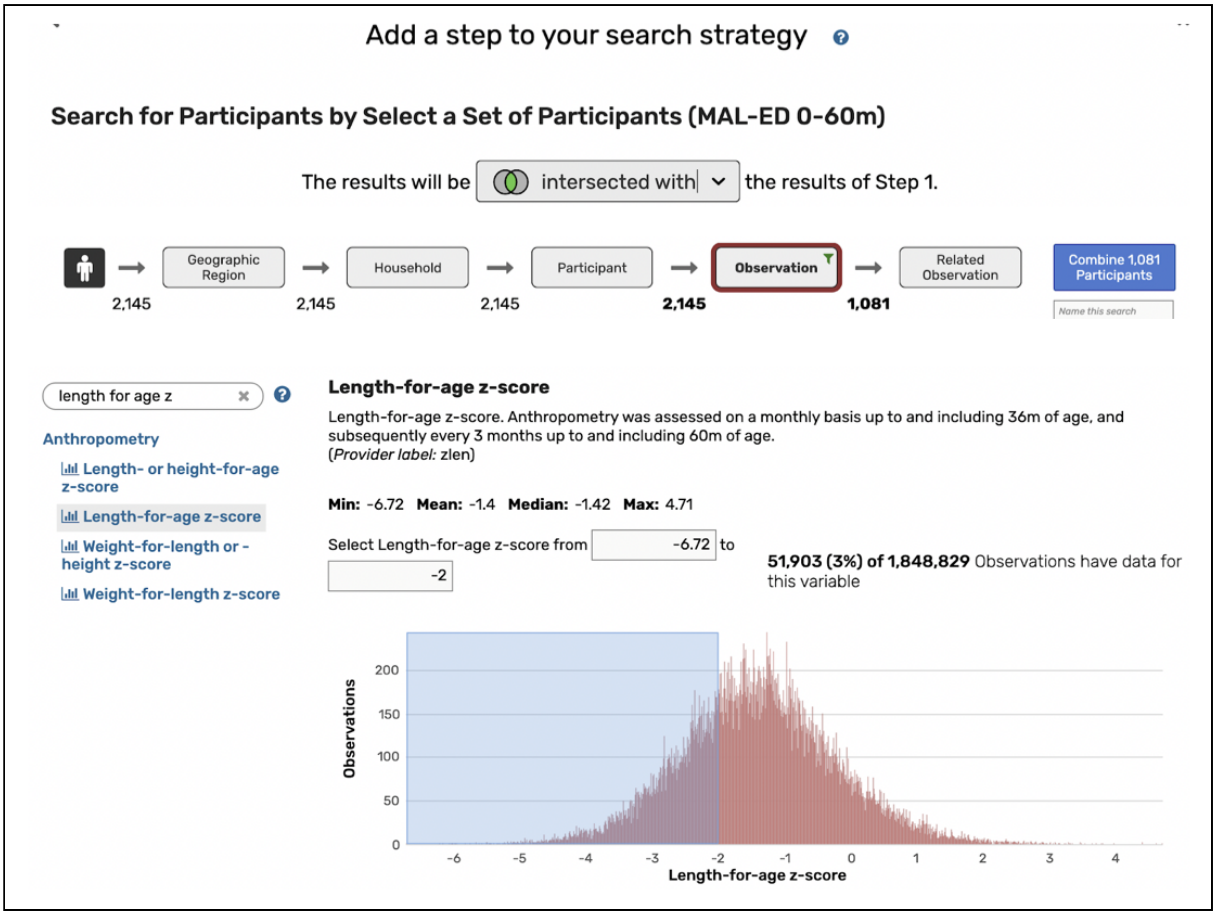
2. In the “Add a step” popup window, click to select “Combine with other MAL-ED 0-60m Participants” and then choose how to combine this new search (search 2) with the original search (search 1). Since we want to identify all participants who had a length-for-age z-score less than -2 at 2 years of age (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.” Finally, choose “A new search” to combine a new subset of MAL-ED participants with the subset of participants identified in the original search (search 1).



3. Click “Select a set of participants (MAL-ED 0-60m)”. The “Add a step” popup window will now contain 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).

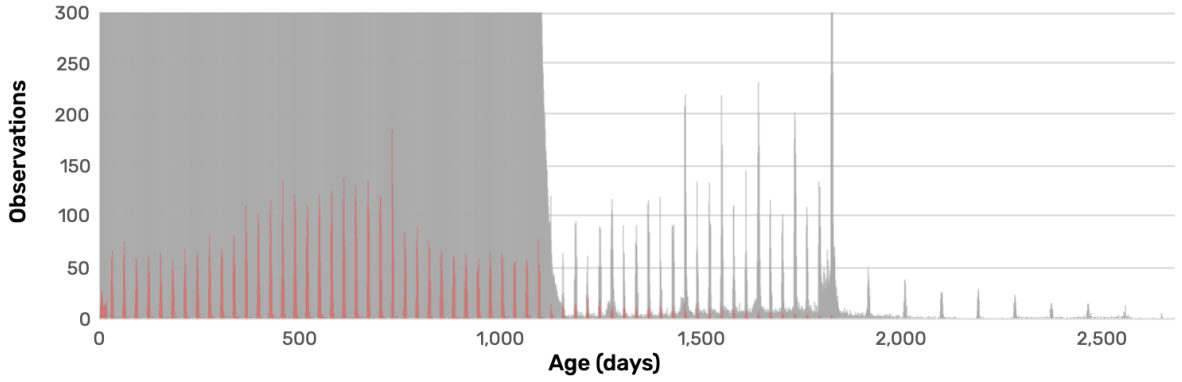


4. For search 2, we want to identify the list of participants who had a length-for-age z-score less than -2 at 2 years of age. The first step in this search is to select participants who had a length-for-age z-score less than -2 at any point during the study. Z-scores are sorted under the “Observations”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



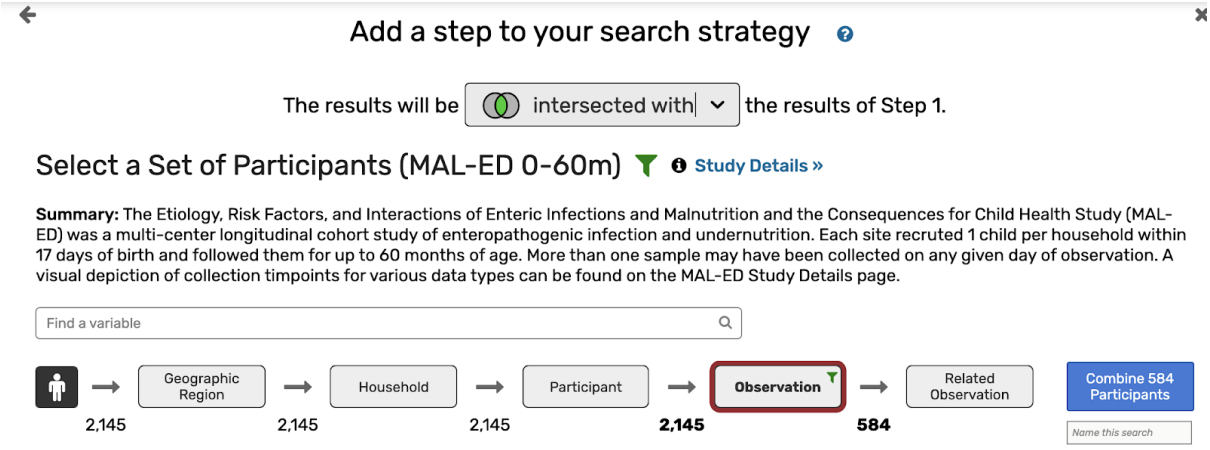
5. We still need to limit the selected length-for-age z-scores to those that were collected when the participant was 2 years of age. Navigate to the "Age (days)" variable in the "Observations" step of the Search Wizard.

6. Use the "Plot settings" under the graph to zoom in on the data; set the y-axis to range from 0 to 300. 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?[4]



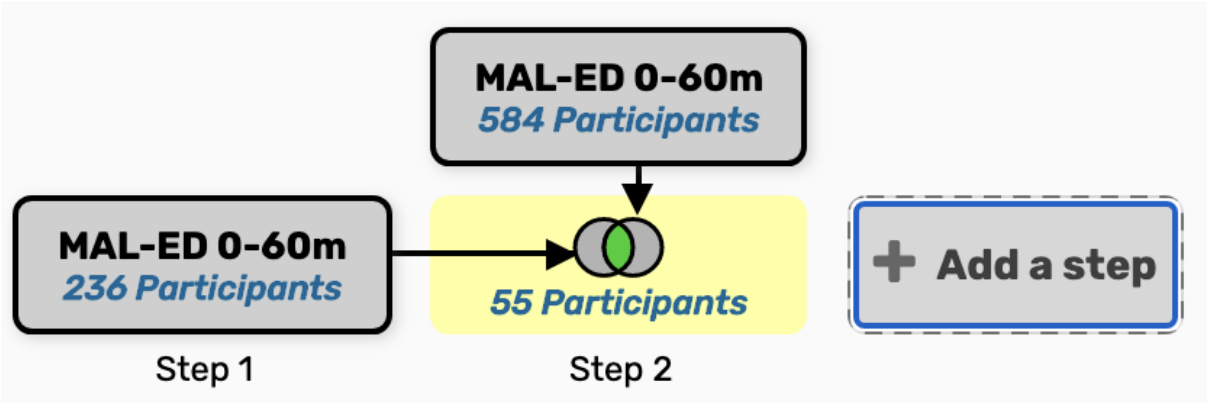
7. Select data that was collected at approximately 2 years of age. There is variation on exactly when the 2 year timepoint of anthropometry data was collected, so be sure to account for this variation by choosing 710 to 750 days for the "Age (days)" range.

8. Search 2 determined that there were 584 MAL-ED participants who had a length-for-age z-score less than -2 at 2 years of age. Click on the blue "Combine 584 Participants" box in the Search Wizard to combine the results of search 2 with search 1.



[4] 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 *Cryptosporidium* within the 14 days before a diarrheal episode at least five days in duration and also had a length-for-age z-score of less than -2 at 2 years of age?[5]



[5] There were 55 participants that were returned in both search 1 and search 2.