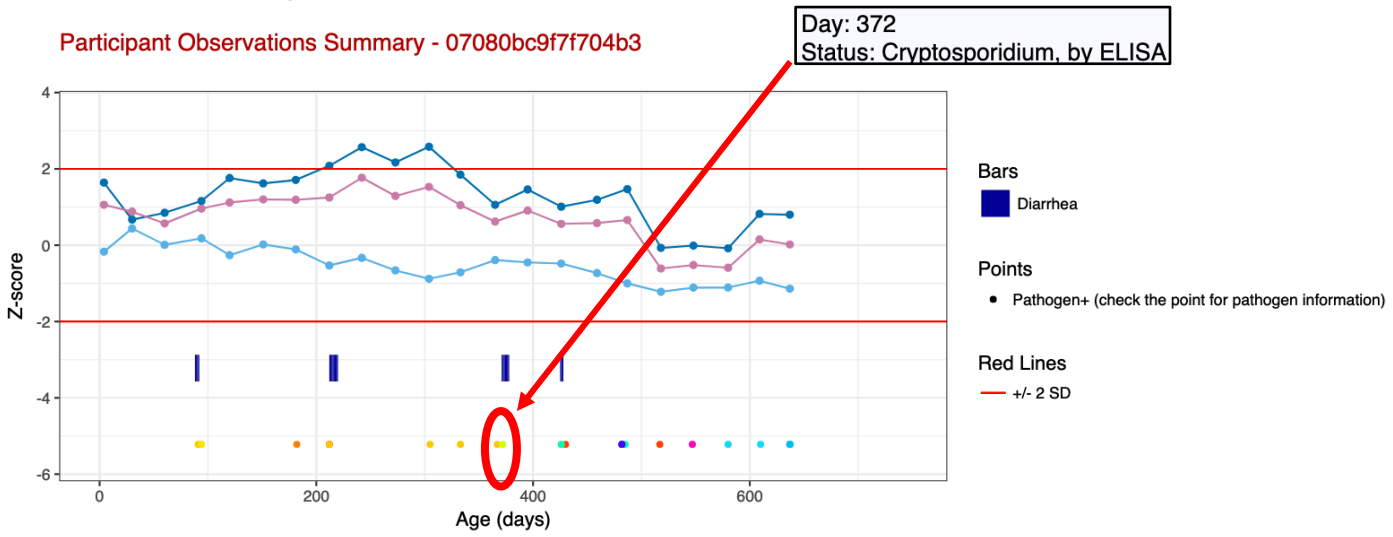
# Building a Search of MAL-ED Data with Related Observations and Revising your Search Strategy

In this exercise, you will build your skills further and learn to do a more advanced search using our related observations function. One of the greatest assets of the MAL-ED data is the extensive 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 of height-for-age z-scores, measured over time for Participant ID 07080bc9f7f704b3.*

**Z-score Observations Graph**




*During follow-up this participant had two episodes of diarrhea that appear to be greater than a few days in length (the blue tick marks on the graph indicate diarrheal episodes; the thickness of each tick mark is proportional to the duration of that diarrheal episode).*

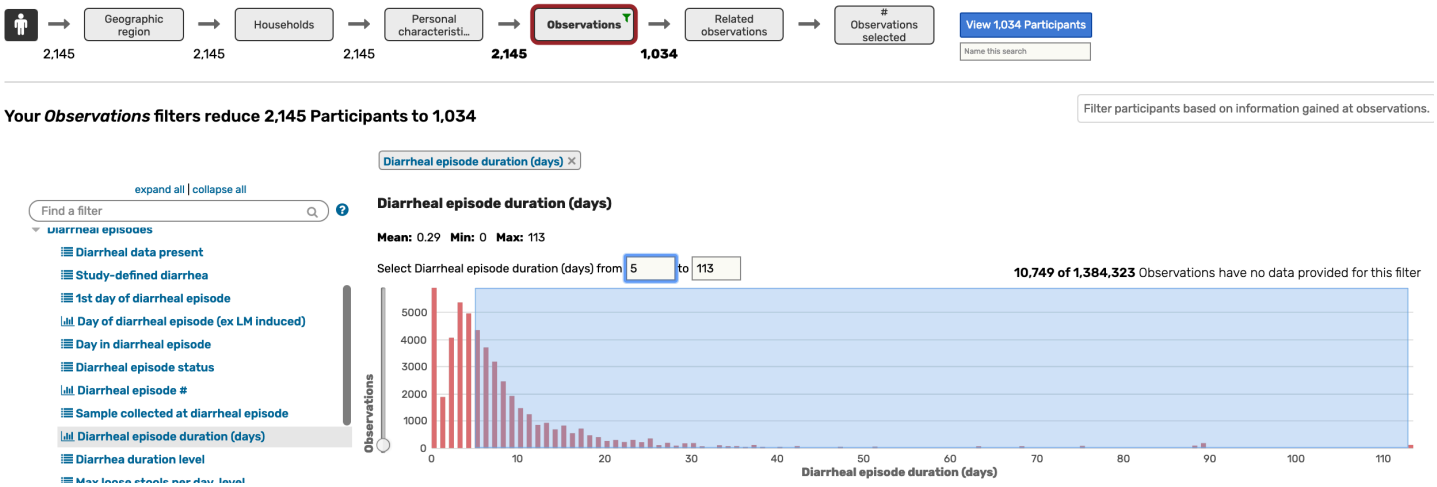
*You can also see that this participant had multiple instances of positive laboratory results (colored dots at the bottom of the graph). One question you might want to ask is,* ***"Did participants tend to test positive for a certain type of pathogen before an episode of diarrhea that was five or more 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 simple.*

1. Start a participant-level search of the MAL-ED data. For this example, 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.

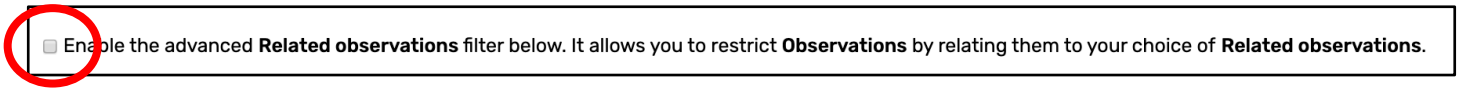



    - Notice that the number of participants has been reduced to 1,034. This means that 1,034 participants had at least one diarrheal episode of at least 5 days of duration at any point in the study. Keep in mind that a participant could have had more than one episode of diarrhea that is at least 5 days of duration.

2. Now, determine how many children tested positive for *Cryptosporidium* within the 2 weeks (14 days) before a diarrheal episode that lasted for at least 5 days. To do this, you would use the 'Related Observations' box in the Search Wizard. The "Related Observations" box is used to narrow the dataset based on the relationship between two variables at two different points in time.

    - Click on "Related Observations" In the Search Wizard.

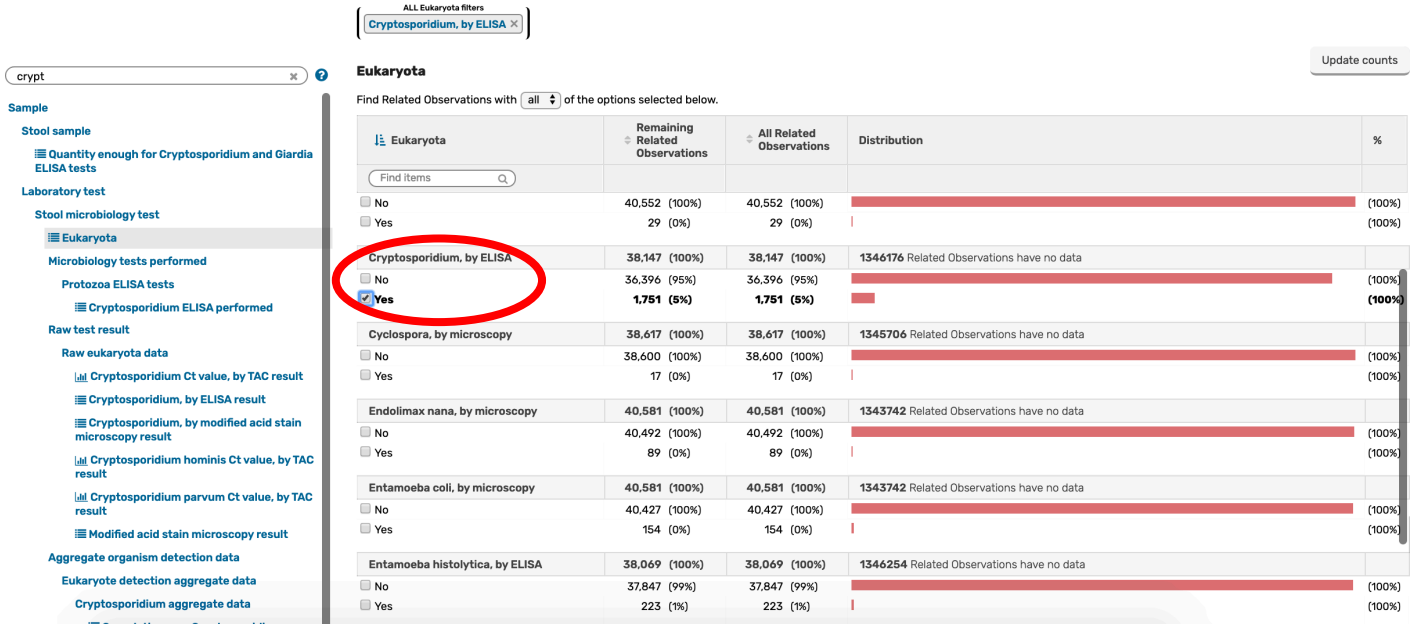

    - Then, click on the small check-box next to the sentences: "Enable the advanced Related observations filter below. It allows you to restrict Observations by relating them to your choice of Related observations."



- How would you complete the 4 boxes in the sentence below if you want to keep the previously selected 5+ day diarrheal observations that occurred 2 weeks after a positive lab test for *Cryptosporidium*?[1] *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

3. Then, click the "Yes" box next to "*Cryptosporidium,* by ELISA" in the multifilter selection table, under the "Eukaryota" category.



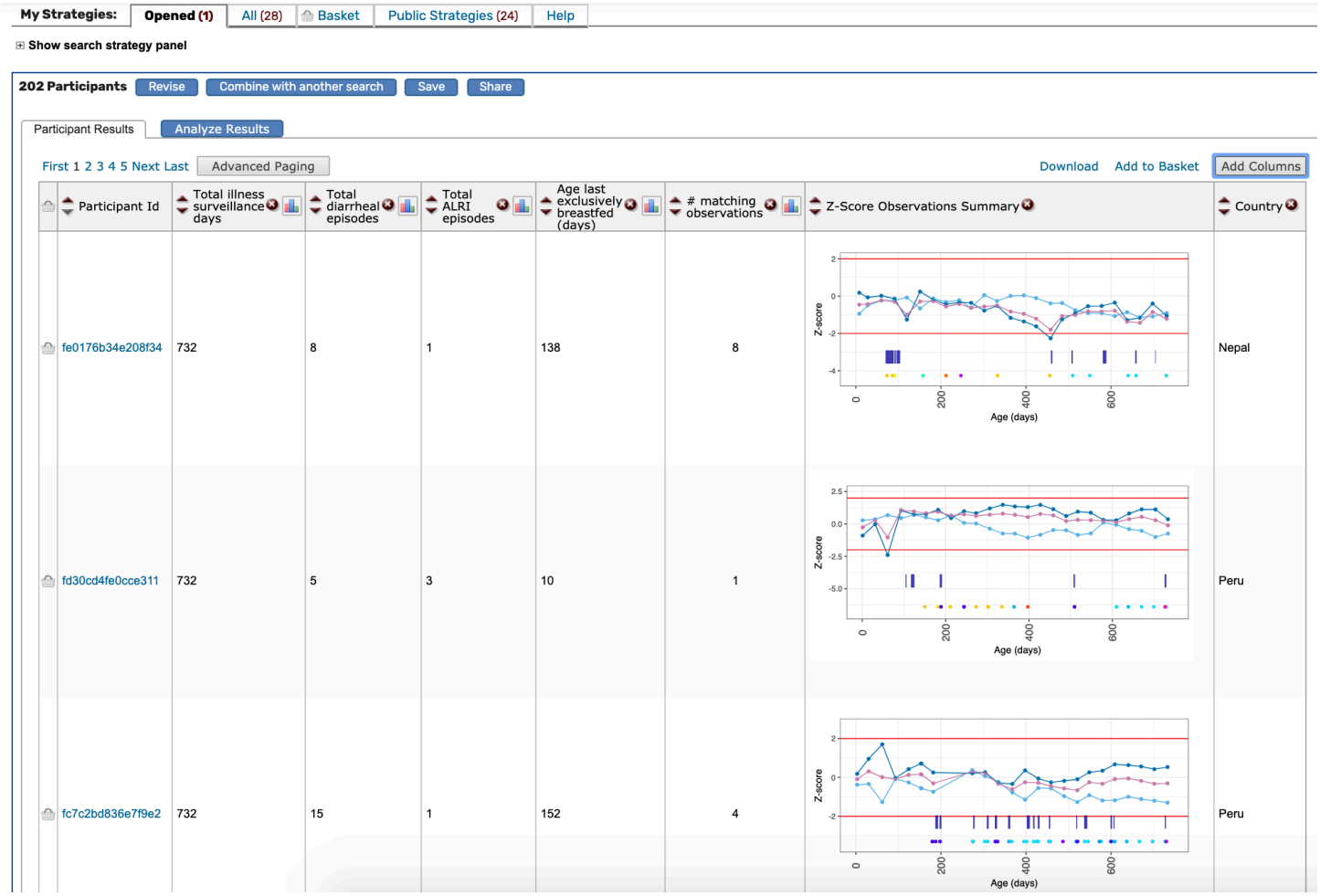

- How does this impact the number of participants that match your search? How many participants have you selected?[2]

1

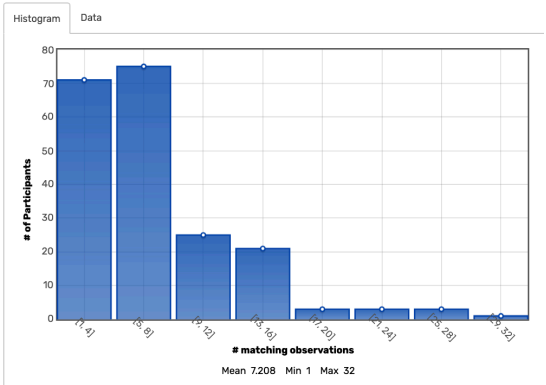
Keep **Observations** that are 0 to 14 days after the **Related observations** specified below

[2] 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.

4. Click on the "View 202 Participants" box in the Search Wizard to see the "Results Table" for these participants. If your default columns 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.




5. "# 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.



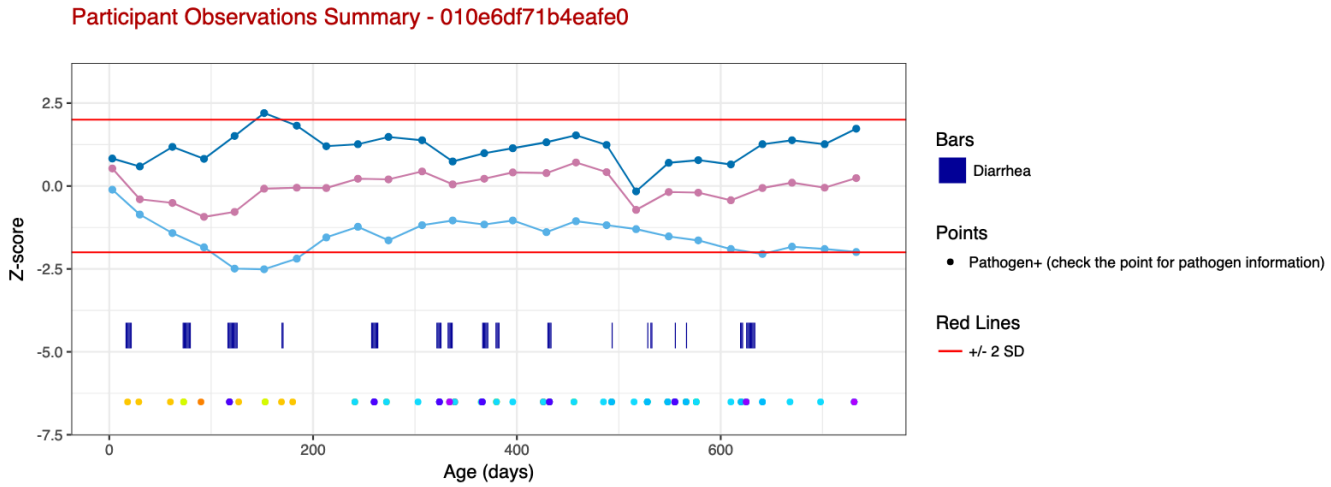

- Notice that some Participants met the search criteria more than once. This means that during the course of the study, some Participants had at least 2 episodes of diarrhea that were 5+ days in duration preceded by a positive lab result for *Cryptosporidium* within the preceding 2 weeks.

- How many participants met the search criteria exactly 1 time?[3] *Hint: Click on the "Histogram" icon on the column header for "# matching observations" and navigate to the "Data" tab.*

[3] 16 participants met the search criteria exactly 1 time.

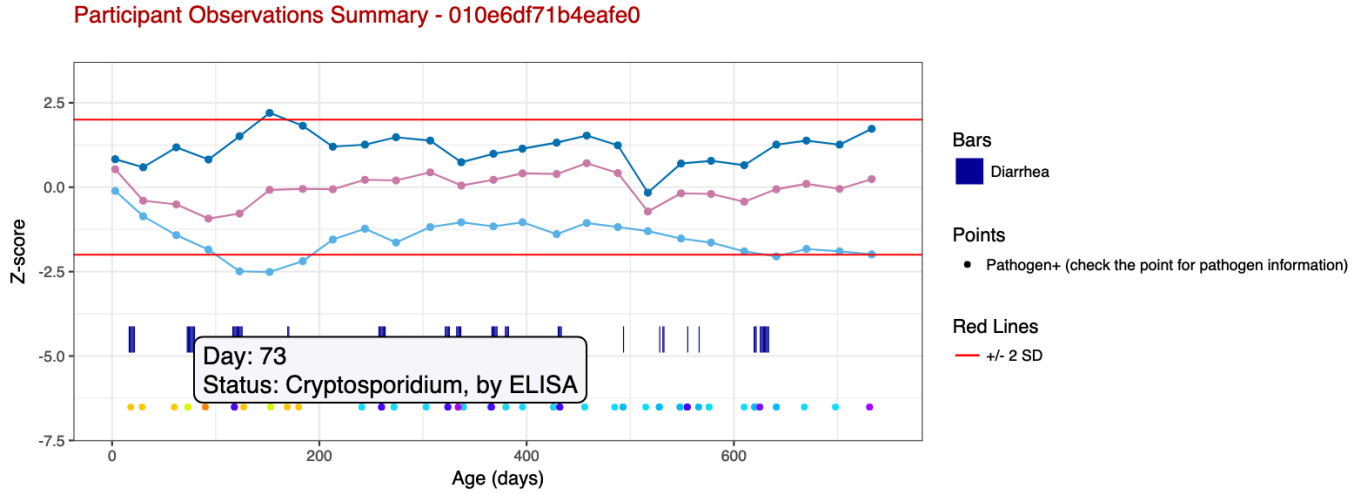
6. Sort the Results Table by "Participant ID" by clicking on the "Participant ID" column header. Participant 010e6df71b4eafe should appear in the first row. Click on 010e6df71b4eafe to bring up data on this specific individual. In the Z-score observations graph, scroll over the colored dots on the bottom of the graph. These colored dots represent positive microbiology test results for different organisms. 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?[4]

**Z-score Observations Graph**




[4] 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.

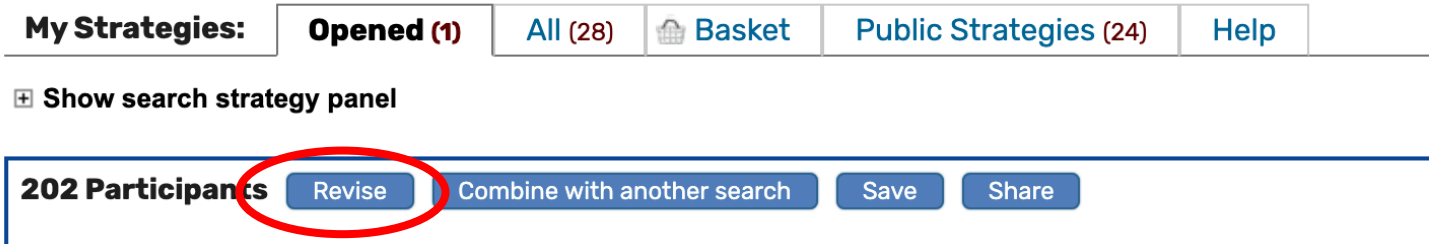
**Z-score Observations Graph**



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 preceded within the previous 14 days by a positive lab result of *Cryptosporidium*.

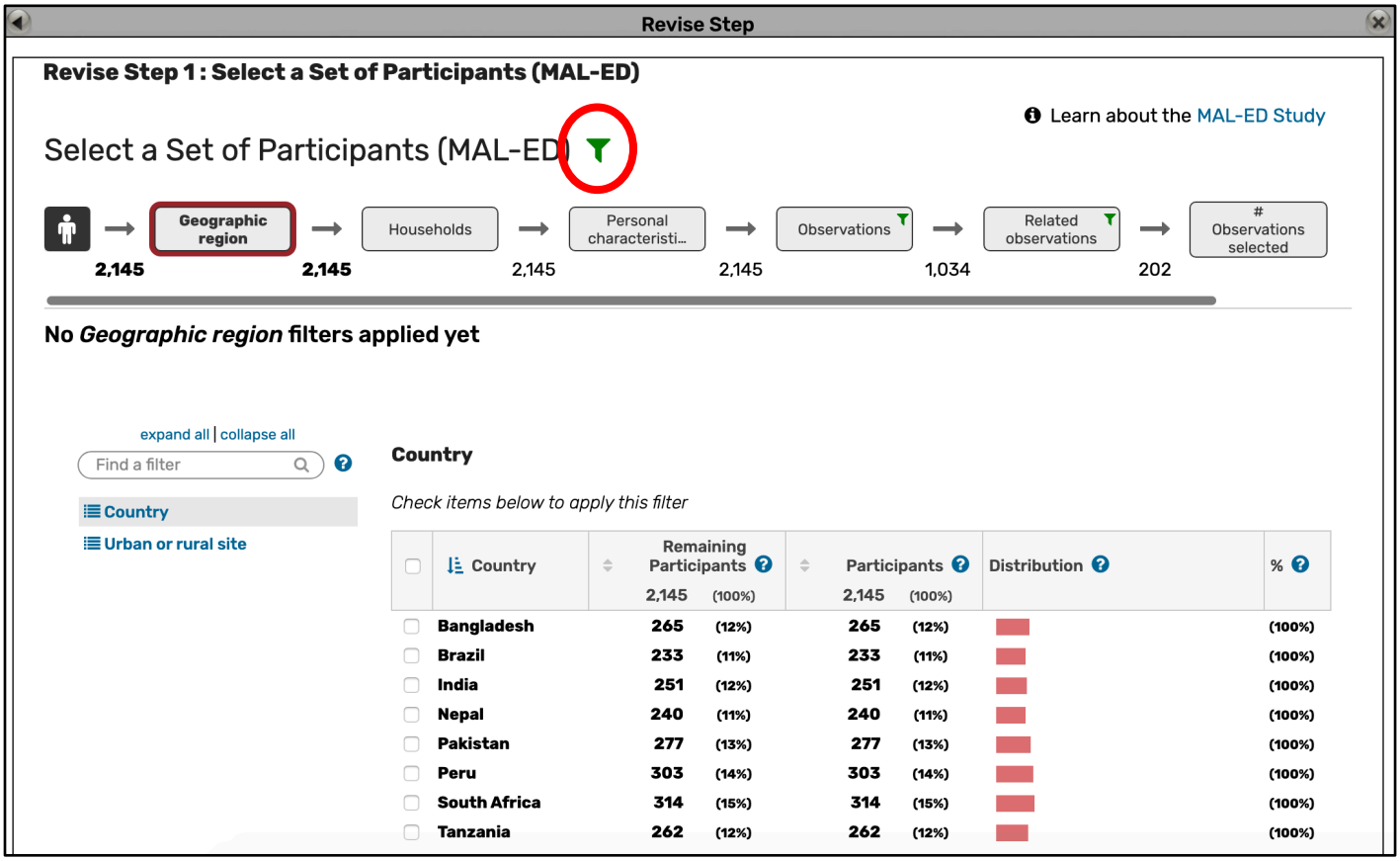
How many participants may have experienced the same 5+ day diarrheal episode, except that they tested positive for *Campylobacter* instead of *Cryptosporidium* in the previous 2 weeks? To do this you can either rerun the entire search and change the microorganism of interest, or, you can simply revise this search as follows:

7. Click on the blue revise button above your Results Table.

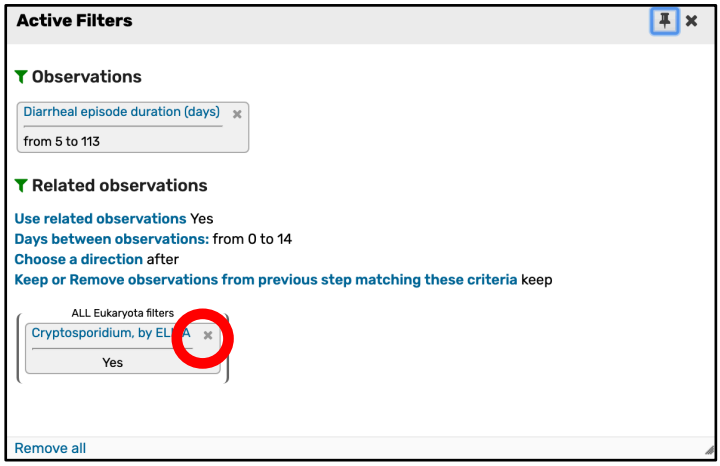



8. Once you click on the revise button, a popup will appear that includes the Search Wizard with your previously selected parameters. Remove the "*Cryptosporidium*, by ELISA" selection criteria:

    - Click on the green "Filter" icon next to "Select a Set of Participants (MAL-ED)" on the top left of the Search Wizard. This will open up a window that lists all of the active filters, which were applied in the previous search.



- Click on the "X" next to the "*Cryptosporidium*, by ELISA" filter to remove this criteria from the search strategy.




9. Now, click on the "Related Observations" box in the Search Wizard and navigate to the filter variables for *Campylobacter* under the "Bacteria" sub-category. Check the "Yes" boxes next to "*Campylobacter*, by ELISA" in the multifilter selection table. Now how many participants have you selected?[5]

10. 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.

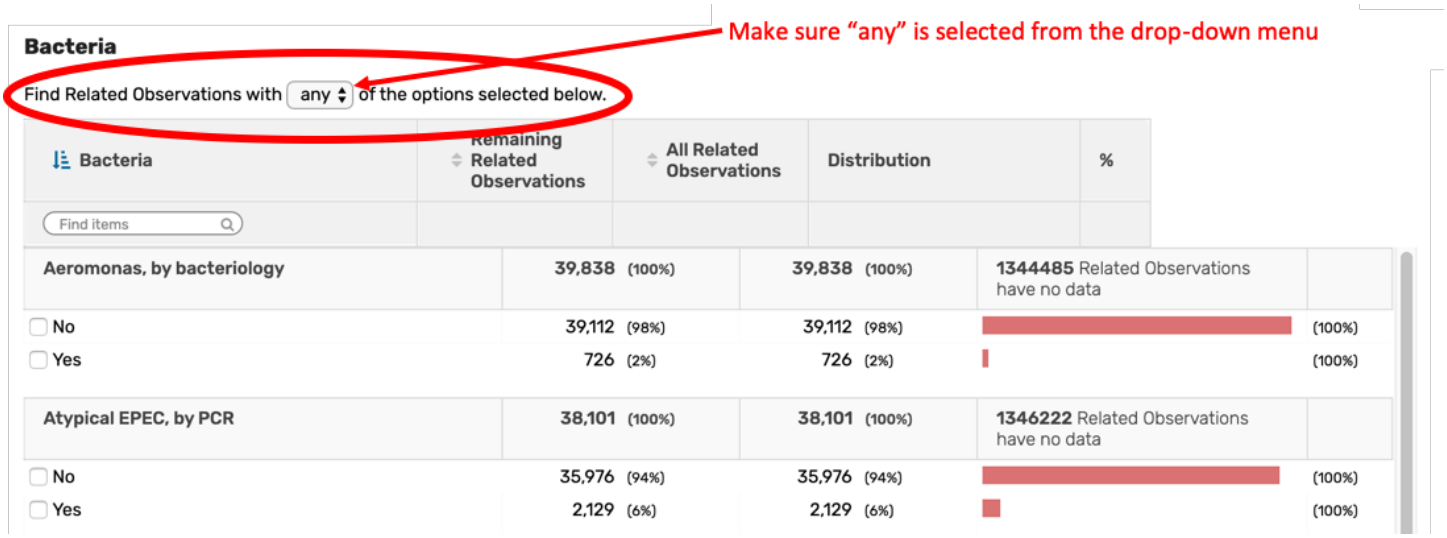
    - What happens to the number of participants that you have selected, and why has this occurred?[6]

    - How would you select participants who tested positive for *Campylobacter* by either methodology? How many participants meet this new search criteria?[7]

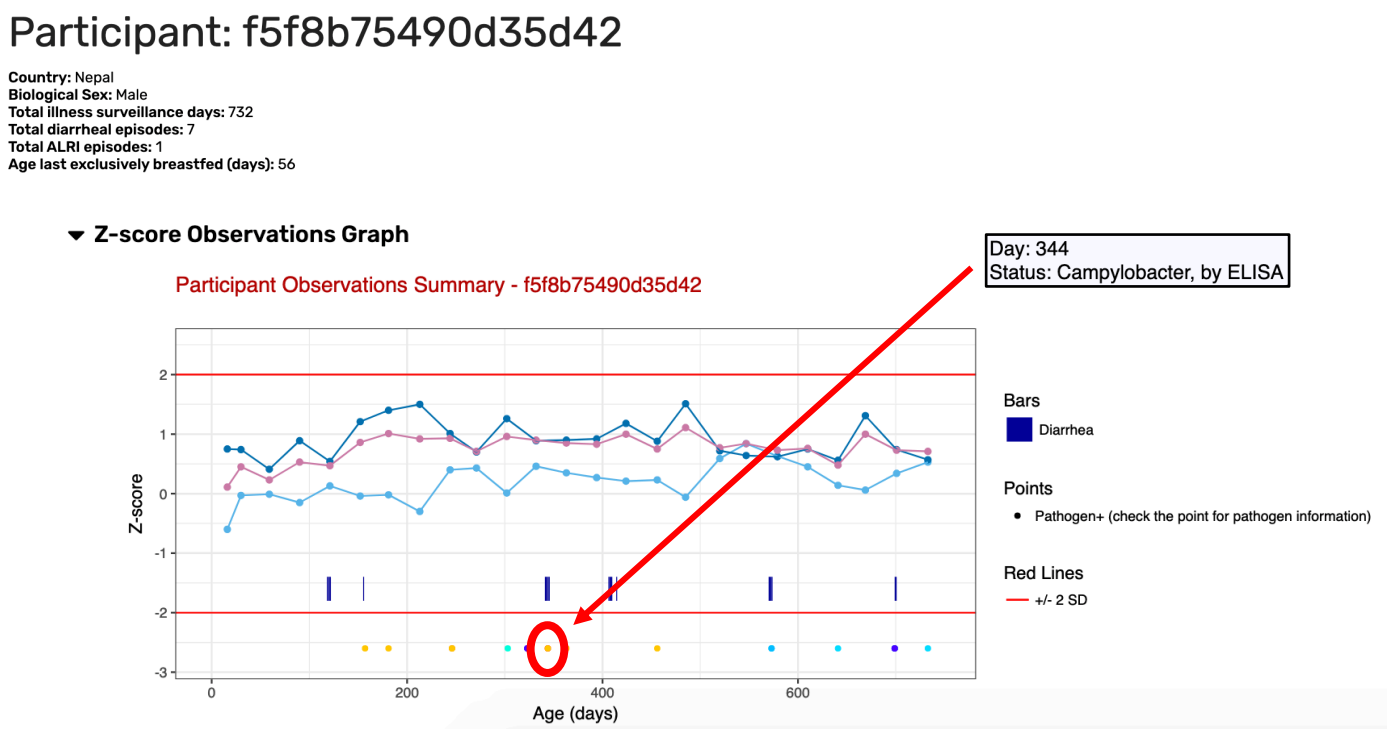
[5] 559 participants had a 5+ day diarrheal episode preceded by a positive *Campylobacter* bacteriology result within the prior 2 weeks.

[6] 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.

[7] 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.



11. Click on the "**View 571 Participants**" 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.




12. 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.