Exercise 3.2: PROVIDE: perform a complex search

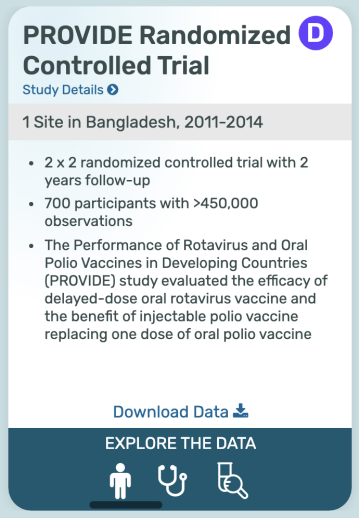
This exercise will give you an opportunity to practice utilizing the search wizard to perform a complex search. You will gain experience using "Related observations" and "Adding a step" to your search strategy.

Within the group of participants who were vaccinated/not vaccinated with Rotarix as specified in the study protocol and who were followed in the study until they reached 2+ years of age, determine:

A. How many times was *Rotavirus* detected by qPCR ***AFTER* the participant was at least 17 weeks of age (the age at which the second dose of Rotarix was given)?**

B. How many of the *Rotavirus* detections in (A) occurred in participants randomized to receive Rotarix?

C. In the set of samples that corresponds to the *Rotavirus* detections in (A), how many samples were also positive for *Norovirus?*



1. What type of search should we execute to determine how many times *Rotavirus* was detected within your population of interest?[1] On the ClinEpiDB.org homepage, click on the appropriate icon on the PROVIDE study card to initiate this search type.

2. Next, we need to start defining our study population of interest. Use "**Rotavirus vaccination per protocol**" to determine which participants received the Rotarix vaccine as specified in the study protocol. Note that the population treated "Per protocol" includes individuals assigned to receive Rotarix oral rotavirus at both 10 and 17 weeks of age and also control individuals who were assigned to NOT receive Rotarix.

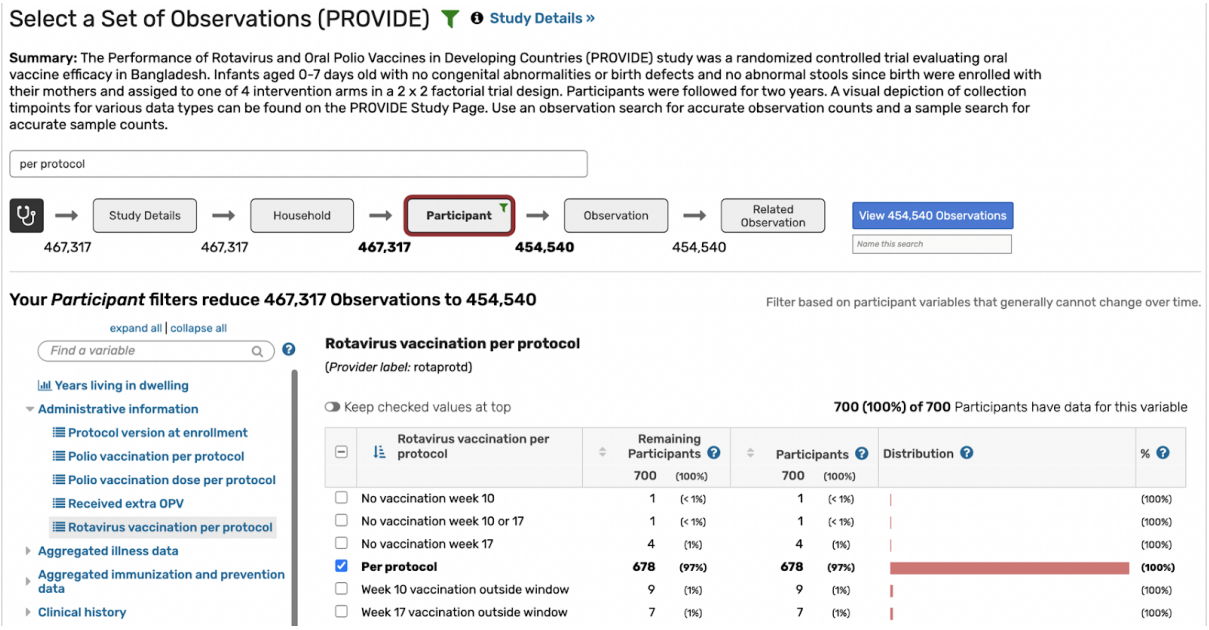
   Remember that you can type text into the "Find a variable" search box ABOVE the search wizard to quickly locate a variable, regardless of which box in the search wizard it has been organized under.

   How many participants were treated "Per protocol"?[2] How many observations were there in these participants?[3]

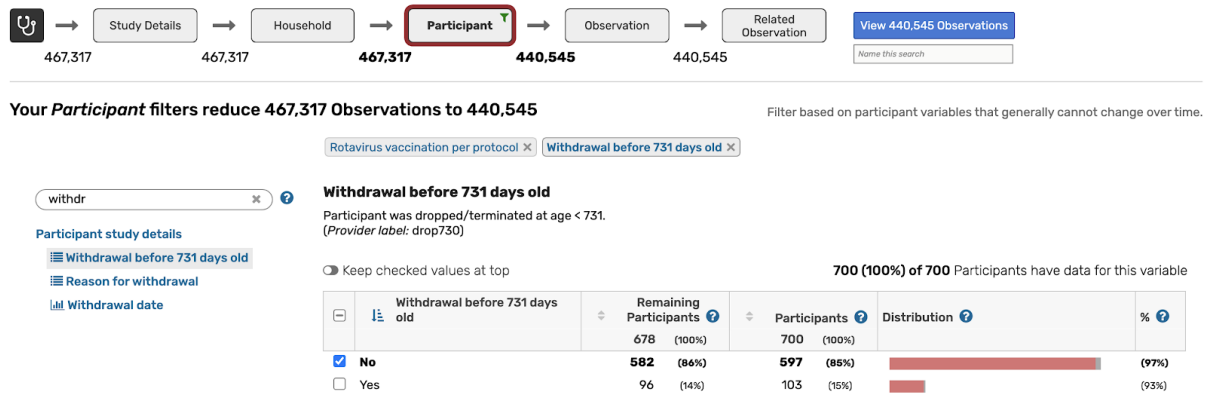
[1] Since we are interested in the total number of times something occurred, we should start an Observation-level search. Remember that since PROVIDE data was collected longitudinally, each participant may have had more than one instance of *Rotavirus* detection.

[2] 678 participants were treated as indicated in the study protocol, meaning that they were each either given the Rotarix oral rotavirus vaccine in accordance with the treatment schedule (if assigned to the treatment arm) or were not vaccinated (if assigned to the control arm).

[3] There were 454,540 observations of these 678 participants.

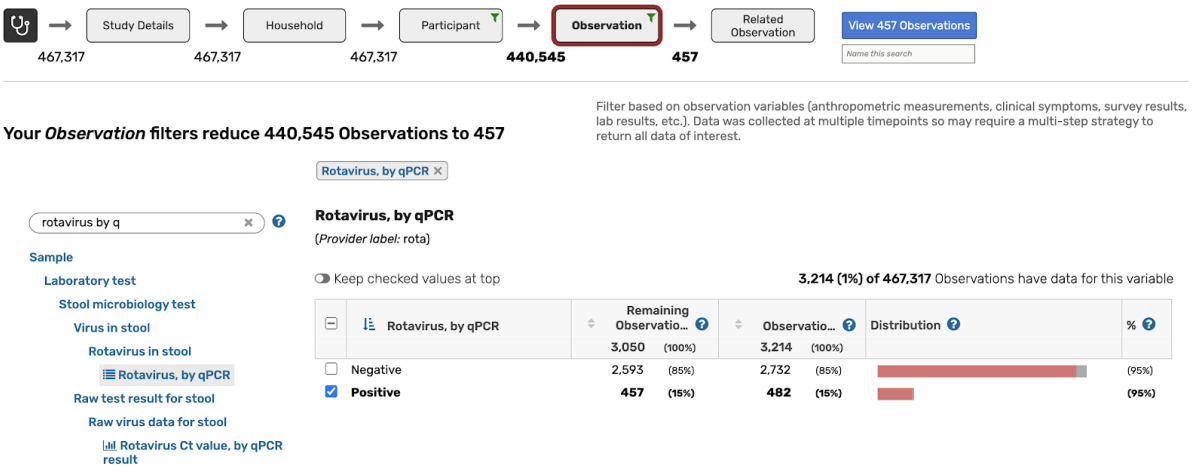


3. Next, further refine your study population of interest, limiting participants to those who were followed in the study until they reached 2+ years of age. Use "**Withdrawal before 731 days old**" to identify the set of participants who were followed until they were at least 2 years of age. Amongst your selected participants who received (or did not receive) the Rotarix vaccine according to the protocol, how many withdrew before 2 years of age (and should be removed from your subset of selected participants)?[4]

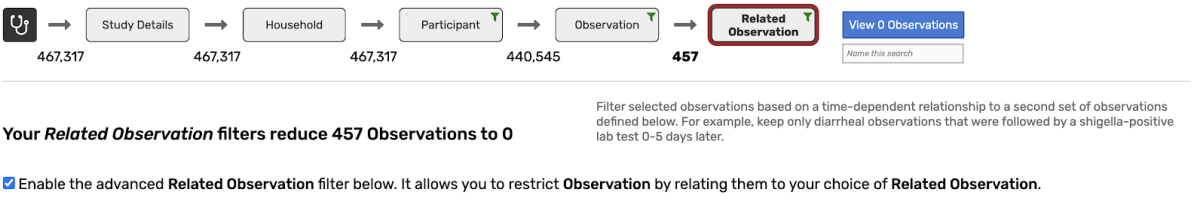


[4]Of the 678 participants who received "Rotavirus vaccination per protocol", 96 withdrew before 2 years of age.

4. Now, let's look at Rotavirus using "**Rotavirus, by qPCR**" to limit our analysis to instances when *Rotavirus* was detected. How many times was *Rotavirus* EVER detected in our population of interest?[5]



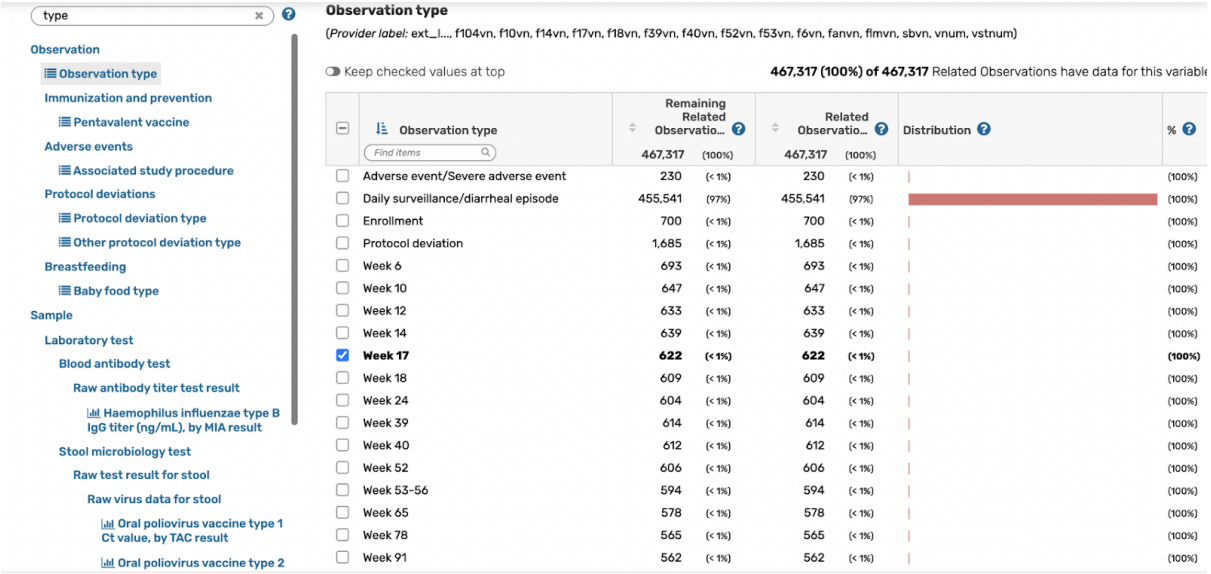
5. Finally, to determine the number of times *Rotavirus* was detected by qPCR ***AFTER*** the participant was 17 weeks of age (the age at which the second dose of Rotarix was given in the treatment arm), we will need to use "Related observations". Click to enable the advanced "Related observation" filter.



6. Remember that the "Related observations" step will require us to (i) complete the sentence indicating which observations to keep or remove and (ii) define the related observations of interest. Let's define the related observations of interest now. We will determine how to complete the sentence next.

   There were 457 observations where *Rotavirus* was EVER detected in our population of interest (***Search 1***). observations. For the related observations search, we want to find all of the observations that occurred in our population of interest ***AFTER*** the participants were 17 weeks old. Use "**Observation type**" for this search (***Search 2***).

[5] *Rotavirus* was detected in our population of interest 457 times.



7. How should we complete the sentence below to properly combine our related observations (*Search 2:* observations occurring post-17 weeks of age) with our observations (*Search 1:* observations where *Rotavirus* was detected)?[6]

Keep / Remove **Observation** that are # to # days Before / After the **Related Observation** specified below

8. You should be left with 385 observations where:
    - Participants were treated according to the study protocol, and either received Rotarix vaccine on schedule or were not vaccinated (if in the control arm).
    - Participants were followed until they were at least 2 years of age.
    - *Rotavirus* was detected AFTER the participant was 17 weeks of age.

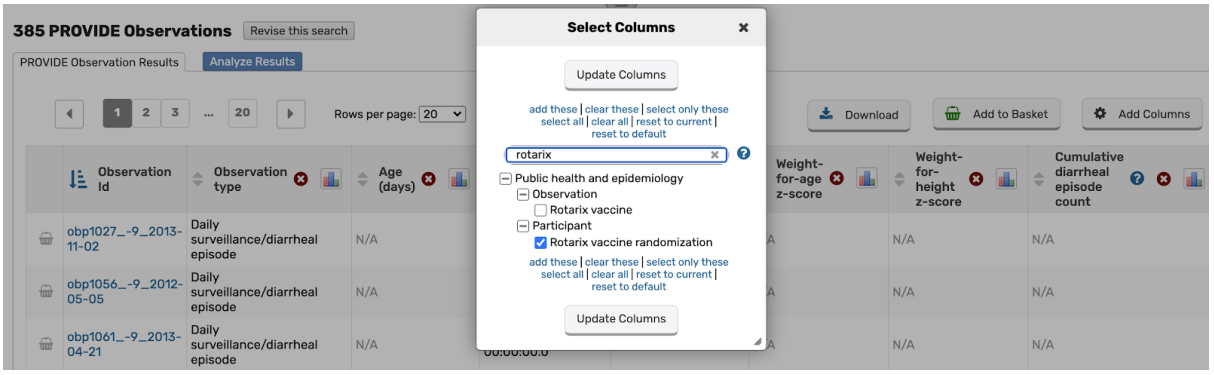
[6] **KEEP** observations that are **0 to 756** days **AFTER** the related observations specified below. We want to KEEP observations (*Search 1,* where *Rotavirus* was detected) that occurred AFTER the participants were 17 weeks old (the related observations, *Search 2*).

Why was the range of days for the related observations set to "0 to 756"?

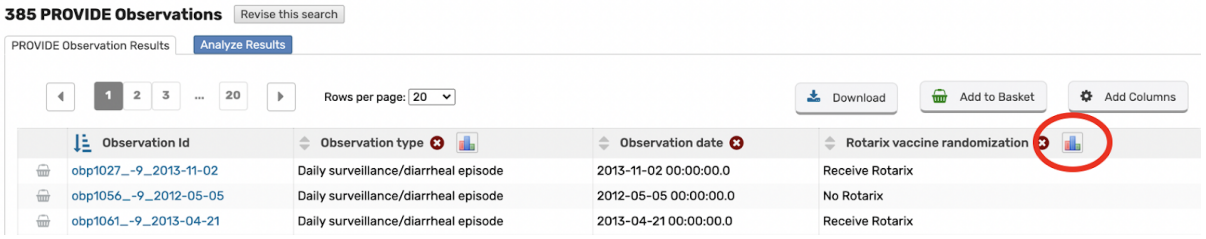
- We are interested in observations that occurred at ANY point in the study after the participant was 17 weeks old, so the minimum was set to 0 (when the participant was 17 weeks old).
- When we look at "**Age (days)**" we can see that the maximum age in the study is 756 days of age. In order to return the full set of observations where the participant was at least 17 weeks of age, we could have completed the range with any number greater than the difference between the maximum age (756) and the age at 17 weeks (119). It's easier to just use 756 instead of calculating that difference.

9. Now that we have limited our observations to those where *Rotavirus* was detected when our population of interest was at least 17 weeks old, we are ready to answer (B): How many of these *Rotavirus* detections occurred in participants randomized to receive Rotarix? Let's click on the blue "View 385 observations" button in the search wizard to get to the results page.

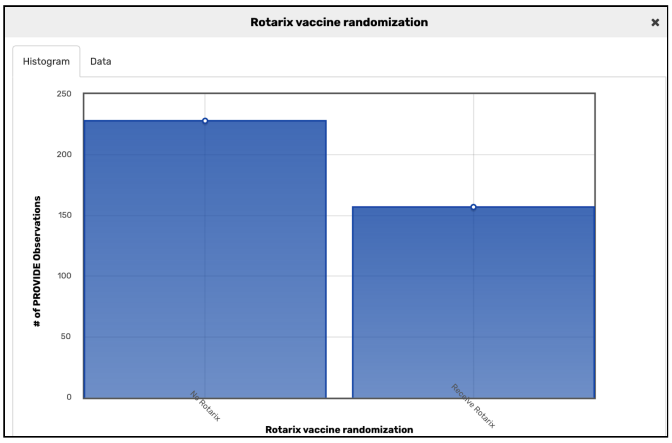
   We are interested in "**Rotarix vaccine randomization**", which will tell us which of these observations occurred in participants who were in the control arm (randomized to receive no rotavirus vaccine) versus the treatment arm (and received the Rotarix vaccine). Click on "Add columns" button on the right-hand side above the Observations table to add this variable.



10. Click on the histogram icon next to the "Rotarix vaccine randomization" column header to view the distribution of observations in participants who received versus did not receive the vaccine. How many observations occurred in participants who received both doses of Rotarix according to the study schedule?[7] *Hint: hover over each of the bars in the histogram, or click on the "Data" tab.*

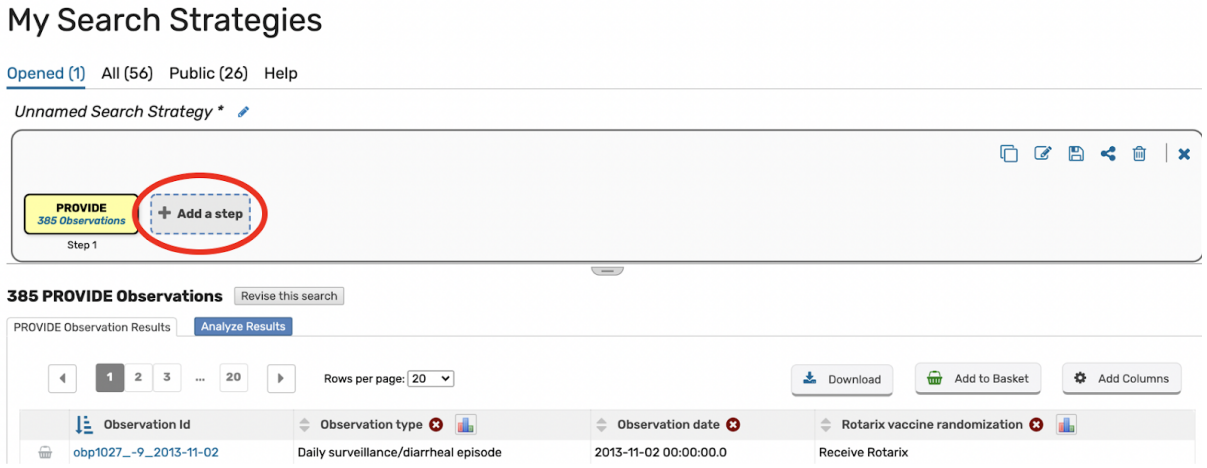


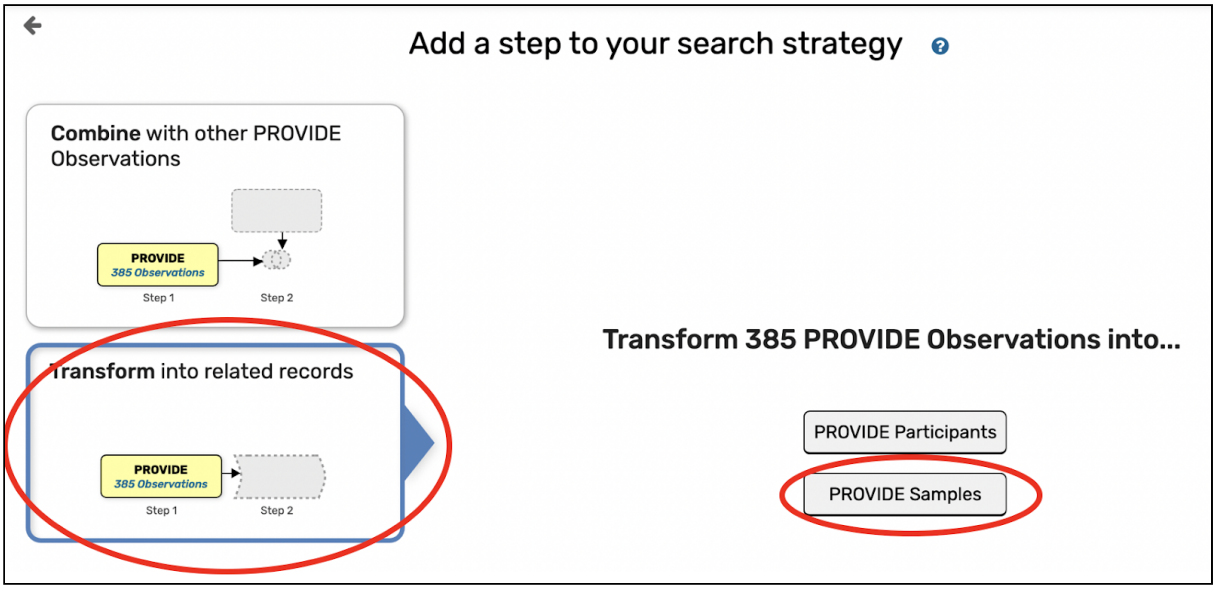
[7] 157 observations occurred in participants who received both doses of Rotarix.

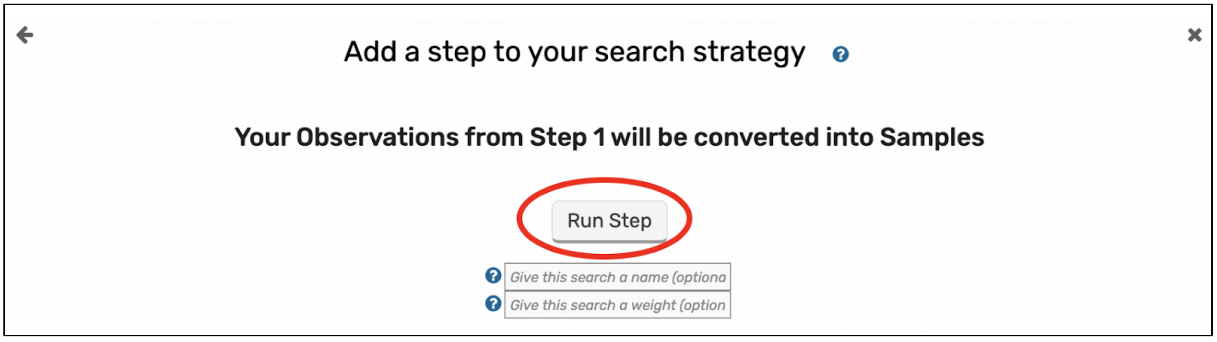


11. Now lets answer question (C): In the set of samples that corresponds to the *Rotavirus* detections in (A), how many samples were also positive for *Norovirus*?

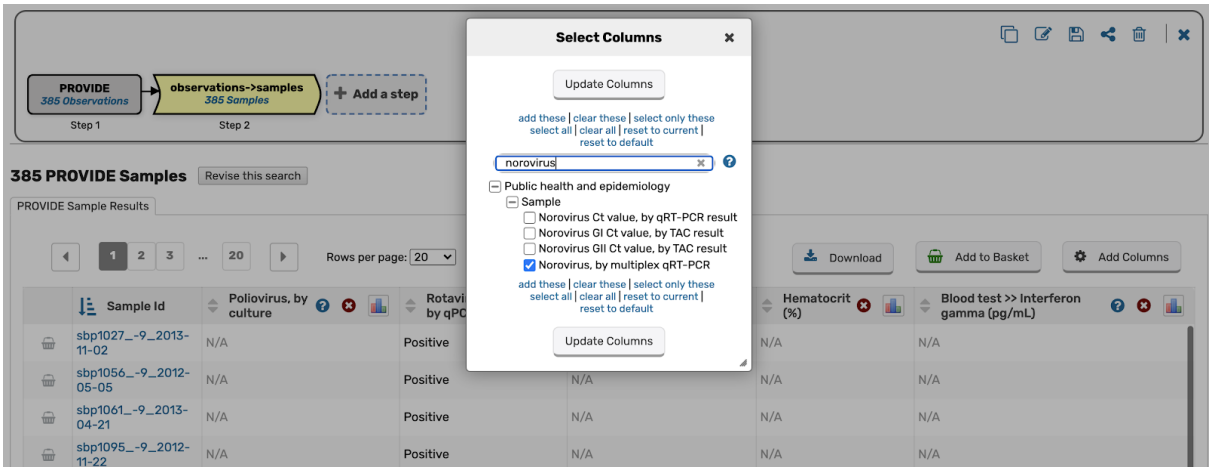
    To answer this question, let's transform our set of observations into the set of samples where *Rotavirus* was detected. Click on "Add a step" in your search strategy, and follow the instructions in the popup windows to transform these 385 observations into their corresponding set of samples.



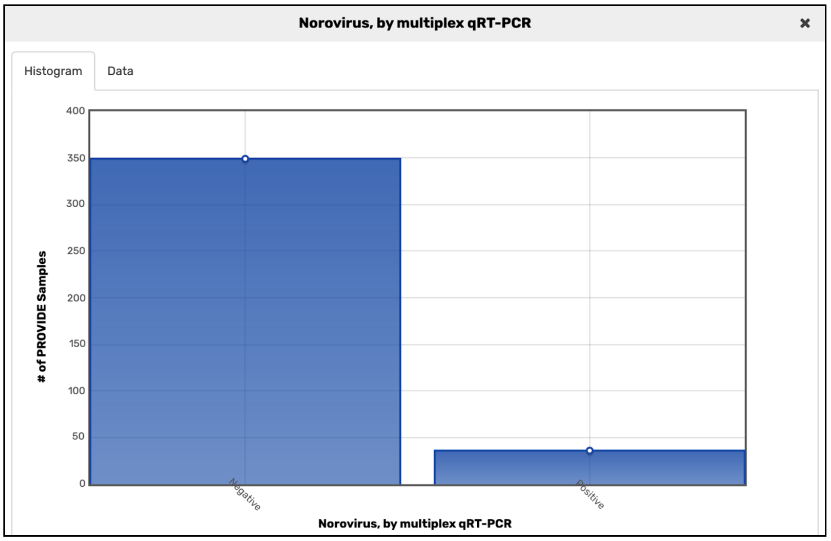




12. We are interested in "**Norovirus, by multiplex qRT-PCR**", which will tell us whether these samples had *Norovirus* detected or not. Click on "Add columns" button on the right-hand side above the Samples table to add this variable.



13. Click on the histogram icon next to the “Norovirus, by multiplex qRT-PCR” column header to view the distribution of samples that tested positive versus negative for *Norovirus*. How many of our samples of interest were also positive for *Norovirus?*[8]



[8] 36 of our samples of interest were also positive for *Norovirus.*