

In GEMS1, how many *Giardia* positive cases had at least one matched control that was *Giardia* negative?

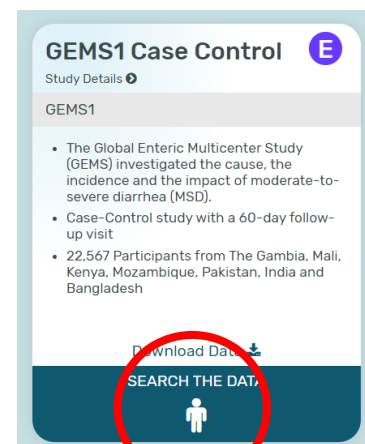
What is the relationship between *Giardia* and MSD? Does this differ for *Cryptosporidium*? What about by age? (30 min)

In this exercise, you will learn to do a more advanced search using our related case/control function. One of the unique assets of the GEMS1 and GEMS1A data is the matched set design and the ability to search these data easily is a valuable feature of the ClinEpiDB resource.

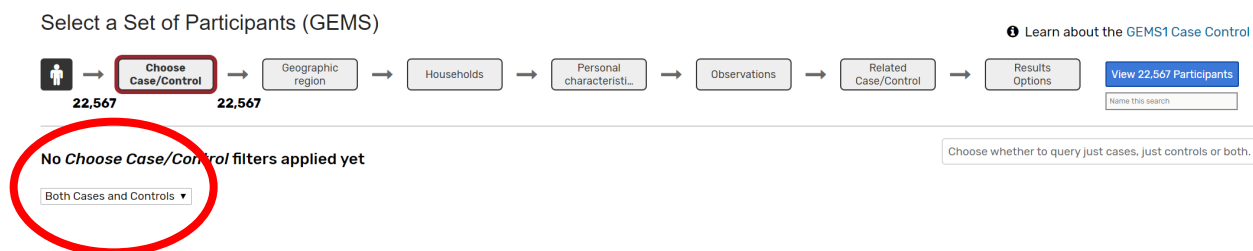
We will start with a very simple related case/control query. For this exercise, to start, we will modify as few filters as possible to keep things simple.

To explore how many *Giardia* positive cases had at least one matched control that was *Giardia* negative, your first step might be to explore the data and see how many Cases tested positive for *Giardia*. Then you would ask the *Giardia* status of those Cases matched controls. *You can do this by using the related Case/Control functionality!*

1. Start this exercise at the GEMS1 Participant level in the search wizard. We describe this functionality as a 'Related Case/Control' search because you are narrowing your dataset based on the relationship between your Case and Control participants. First select Cases who tested positive for *Giardia*.



2. Navigate to the 'Choose Case/Control' box in the Search Wizard. Select 'Cases'



3. Next, select from Observations that *Giardia* positive. This can be found in the 'Stool microbiology tests' filter in observations. You can navigate to this through the filter hierarchy on the left or by typing "Stool microbiology" into the search box.

Select a Set of Participants (GEMS) ▼

Learn about the GEMS1 Case Control



Your **Observations** filters reduce 9,439 Participants to 1,786

Filter participants based on information gained at observations.

expand all | collapse all

Find a filter

ALL Eukaryota filters
Giardia, by ELISA x

Update counts

Eukaryota

Find Observations with all of the options selected below.

	Remaining Observations	All Observations	Distribution	%
Eukaryota				
Cryptosporidium, by ELISA	1,786 (100%)	22,565 (100%)	38,393 Observations have no data	
<input type="checkbox"/> No	1,552 (87%)	20,599 (91%)		(8%)
<input type="checkbox"/> Yes	234 (13%)	1,966 (9%)		(12%)
Entamoeba histolytica, by ELISA	1,786 (100%)	22,565 (100%)	38,393 Observations have no data	
<input type="checkbox"/> No	1,735 (97%)	21,987 (97%)		(8%)
<input type="checkbox"/> Yes	51 (3%)	578 (3%)		(9%)
Giardia, by ELISA	9,438 (100%)	22,565 (100%)	38,393 Observations have no data	
<input type="checkbox"/> No	7,652 (81%)	17,310 (77%)		(44%)
<input checked="" type="checkbox"/> Yes	1,786 (19%)	5,255 (23%)		(34%)

You should notice that the number of participants reduced to 1,786. This means that 1,786 Cases had a positive test for *Giardia*.

- Now, ask that of these *Giardia* positive Cases, how many had at least one matched Control that was *Giardia* negative. To do this, you would use the 'Related Case/Control' box in the query wizard.
- Click on 'Related Case/Control'

Select a Set of Participants (GEMS) ▼

Learn about the GEMS1 Case Control



You can restrict the participants you have previously selected by removing (or keeping) participants based on differences or similarities between matched cases and controls.

☒ Enable the advanced **Related Case/Control** filter below. It allows you to restrict Participants using information about their related case or control.

Remove Participants based on your choice of Related Case/Control Participants below

- Then, click on the small check-box next to the sentence 'Enable the Related observations filter below. Enabling this option will allow you to restrict Observations by relating them to your choice of Related observations.'

Think carefully about what we want to do. We want to keep the selected Cases that have a Control that is *Giardia* negative.

Complete the sentence like this:

Keep ▼ Participants based on your choice of Related Case/Control Participants below

Remove

Keep

7. Then, select *Giardia* 'No' from the Stool microbiology test, Eukaryota filter.

Keep ▾ Participants based on your choice of Related Case/Control Participants below

ALL Eukaryota filters
(Giardia, by ELISA X)

Eukary

Sample

Laboratory test

Stool microbiology test

Eukaryota

Raw test result

Raw eukaryota data

Ancylostoma, by TAC result

Ascaris, by TAC result

Cryptosporidium, by TAC result

Cryptosporidium hominis, by TAC result

Cryptosporidium parvum, by TAC result

Cyclospora, by TAC result

Enterocytozoon bienersi, by TAC result

Eukaryota

Find Observations with all ▾ of the options selected below.

	Remaining Observations	All Observations	Distribution	%
Eukaryota				
Cryptosporidium, by ELISA	13,127 (100%)	22,565 (100%)	38,393 Observations have no data	
No	12,284 (94%)	20,599 (91%)		(60%)
Yes	843 (6%)	1,966 (9%)		(43%)
Entamoeba histolytica, by ELISA	13,127 (100%)	22,565 (100%)	38,393 Observations have no data	
No	12,828 (98%)	21,987 (97%)		(58%)
Yes	299 (2%)	578 (3%)		(52%)
Giardia, by ELISA	13,127 (100%)	22,565 (100%)	38,393 Observations have no data	
No	9,658 (74%)	17,310 (77%)		(56%)
Yes	3,469 (26%)	5,255 (23%)		(66%)

Update counts

How does this impact the number of participants that match your search? There should only be 1,259 participants remaining.

Select a Set of Participants (GEMS) ▼



No Related Case/Control filters applied yet

You can restrict the participants you have previously selected by removing (or keeping) participants based on differences or similarities between matched cases and controls.

8. Click on the 'View 1,259 Participants' box to see the table listing of these Participants. Click on the small histogram icons next to the column names to see the distribution of characteristics in this sample.

1259 Participants

Revise Combine with another search Save Share

Participant Results Analyze Results

First 1 2 3 4 5 Next Last Advanced Paging

Download Add to Basket Add Columns

Participant Id	Case or Control Child	Sex	Country
1010001181	Case	Female	The Gambia
1010016241	Case	Male	The Gambia
1010038821	Case	Male	The Gambia
1010040741	Case	Male	The Gambia
1010044131	Case	Female	The Gambia

To change the default columns listed in the result click on the 'Add Columns' button located at the top right corner of your Participant result tab and add the desired columns. Try adding 'Child given untreated drinking water' to the columns.

9. Save this search by clicking on the 'Save' button above the search results. You will need to login to save.

My Strategies: **Opened (1)** All (1) Basket Public Strategies (2) Help

Show search strategy panel

1259 Participants Revise Combine with another search Save Share

Participant Results Analyze Results

First 1 2 3 4 5 Next Last Advanced Paging

Participant Id	Case or Control Child	Sex
1010001181	Case	Female
1010016241	Case	Male
1010038821	Case	Male

Revising Your Search

From your previous search, you now know that there were 1259 Cases in the GEMS study that were positive for *Giardia* had at least one matched control that was negative for *Giardia*. Maybe now you would like to ask, what is the relationship between *Giardia* and moderate-to-severe diarrhea in the first two years of life? For matched case-control studies the most robust analysis of associations is to be done using a matched pairs analysis. Our apps are for exploration and hypothesis generation purposes only and a matched pairs analysis is not available at this time. However, you can generate a simple contingency table to look at associations.

To look at the relationship between *Giardia* and moderate-to-severe diarrhea for your whole sample, rather than just the 1259 participants you just returned, you will want to revise your search to remove any filters you've applied. To do this click the revise button as follows:

1259 Participants Revise Combine with another search

Participant Results Analyze Results

First 1 2 3 4 5 Next Last Advanced Paging

Participant Id	Country	Sex	Age at Eligibility Form
GCCP_1010303481	The Gambia	Male	12
GCCP_1010411311	The Gambia	Female	23

1. Click on the blue revise button above your results table.
2. Once you click on the revise button, a popup will appear that includes the search wizard with your previously selected parameters. To remove all previously applied filters, simply click on the green 'filter' icon anywhere in the pop-up. Another pop-up will open that lists all of your active filters.



Revise Step 1: Select a Set of Participants (GEMS)

Select a Set of Participants (GEMS) [Learn about the GEMS1 Case Control](#)

22,567 → Choose Case/Control 9,439 → **Geographic region** 9,439 → Households 9,439 → Personal characteristics 9,439 → Observations 1,786 → Related Case/Control 1,259 → Results Options [View 1,259 Participants](#)

No Geographic region filters applied yet

Country

Check items below to apply this filter

Country	Remaining Participants	Participants	Distribution	%
Bangladesh	1,394 (15%)	3,859 (17%)	<div></div>	(36%)
India	1,568 (17%)	3,502 (16%)	<div></div>	(44%)
Kenya	1,476 (16%)	3,359 (15%)	<div></div>	(44%)
Mali	2,033 (22%)	4,097 (18%)	<div></div>	(50%)

Click on 'Remove all' to remove all of the active filters then close the 'Active Filters' box with the 'x' in the top right corner.

Revise Step

(GEMS)

Active Filters

- Choose Case/Control
 - Cases or Controls Cases
- Observations
 - ALL Eukaryota filters
 - Giardia, by ELISA x
 - Yes
- Related Case/Control
 - Compare Cases / Controls Yes
 - Remove all

Remaining Participants 9,439 (100%) Participants 22,567 (100%) Distribution

3. Finally, click 'View 22,567 participants' to get back to the full GEMS1 data set.

Revise Step 1: Select a Set of Participants (GEMS)

Select a Set of Participants (GEMS) [Learn about the GEMS1 Case Control](#)

22,567 → Choose Case/Control 22,567 → **Geographic region** 22,567 → Households 22,567 → Personal characteristics 22,567 → Observations 22,567 → Related Case/Control 22,567 → Results Options [View 22,567 Participants](#)

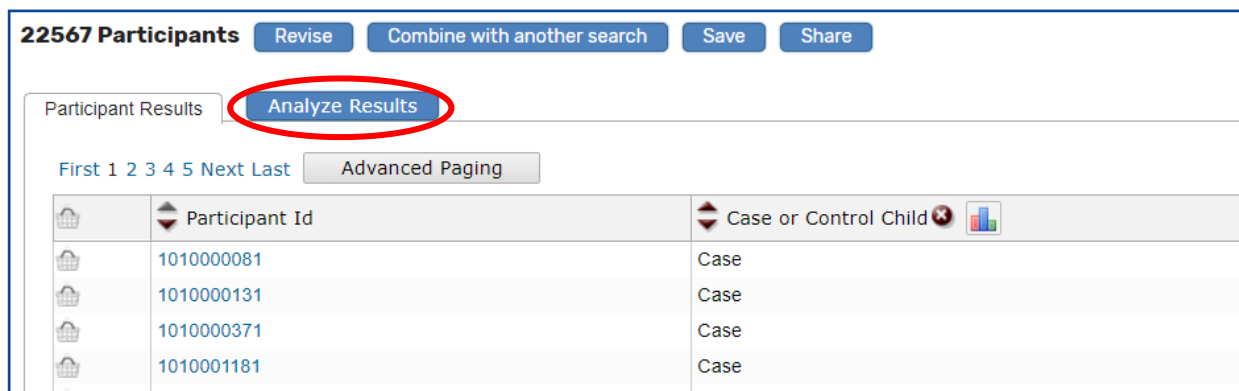
No Geographic region filters applied yet

Country

Check items below to apply this filter

Country	Remaining Participants	Participants	Distribution	%
Bangladesh	3,859 (17%)	3,859 (17%)	<div></div>	(100%)

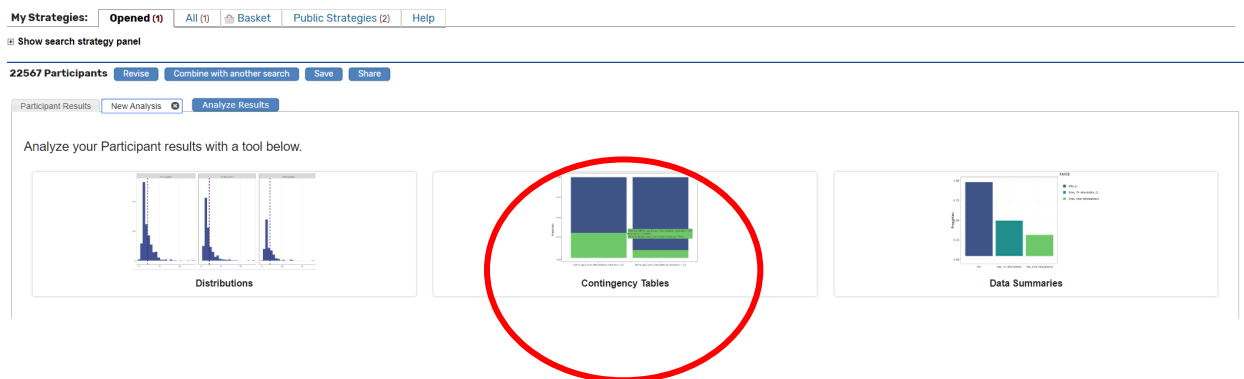
4. To analyze the association between MSD and Giardia, two binary outcomes, you want to run a contingency table. Click on the blue tab next to the Participant Results tab in your data table that says “Analyze Results”



The screenshot shows the top of the Shiny app interface. At the top, it says "22567 Participants" followed by buttons for "Revise", "Combine with another search", "Save", and "Share". Below this is a tab bar with "Participant Results" and "Analyze Results". The "Analyze Results" tab is highlighted with a red circle. Below the tabs, there is a pagination bar with "First 1 2 3 4 5 Next Last" and an "Advanced Paging" button. Below that is a table with two columns: "Participant Id" and "Case or Control Child". The table shows four rows of data, all labeled "Case".

Participant Id	Case or Control Child
1010000081	Case
1010000131	Case
1010000371	Case
1010001181	Case

5. Initiate the app by clicking on the box labeled “Contingency Tables”



The screenshot shows the Shiny app interface with the "Analyze Results" tab selected. Below the tabs, there is a section titled "Analyze your Participant results with a tool below." which contains three boxes: "Distributions", "Contingency Tables", and "Data Summaries". The "Contingency Tables" box is highlighted with a red circle. Above the boxes, there is a "My Strategies" section with "Opened (1)" and buttons for "All (1)", "Basket", "Public Strategies (2)", and "Help". Below this is a "Show search strategy panel" button. At the top of the app, there are buttons for "Revise", "Combine with another search", "Save", and "Share".

This Shiny app allows you to look at the association between any two variables that are found in the dataset. Notice that there are drop-down menus that allow you to choose what variables you are comparing. Variable 1 defines the rows that represent subsets of your population and Variable 2 defines the columns.

7. Examine the default characteristics of the app.

Contingency Tables

Plot Parameters

Summary Statistics

Contingency Tables

Plot Grid

Individual Plots

Help

Variable 1

60 day follow-up conducted

are / is

ever

Yes

Variable 2

Selected Participants

are / is

-Selected Items Will Appear Here-

Facet Plot (1)

All possible

facets for

Selected Participants

Facet Plot (2)

None

As you can see, in the large right panel you have drop down menus that allow you to select which variables you wish to compare. By default, for Variable 1, which defines the rows of the 2x2 table, 60 day follow-up conducted is selected. For variable 2, which defines the columns the participants you have selected (all of them in this case since you have removed all filters) are selected.

8. For this analysis start by selecting 'Case/Control' for Variable 1 and 'Giardia, by ELISA' for Variable 2.

Variable 1

Case or Control Child

Giardia Assemblage B, by TAC result

Giardia, by TAC result

Isospora, by TAC result

Necator, by TAC result

Strongyloides, by TAC result

Trichuris, by TAC result

Virus

Bacteria

Eukaryota

Giardia

Giardia, by ELISA

Entamoeba histolytica

Cryptosporidium

Observation

Household

Variable 2

Case or Control Child

Case

Participant

Case or Control Child

Participant study details

Feeding

Corresponding case

Age, corresponding case (months)

Sex, corresponding case

Administrative information

Child ID of matched case

Study center

Child ID

Case ID

All forms complete

All lab forms complete

All members of the Good Pair have all for

Variable 1

Giardia, by ELISA ▼

are / is

Yes ▼

Variable 2

Case or Control Child ▼

are / is

Case ▼

9. Next, click on the 'Contingency Tables' tab from the left panel. You can see that the numbers within the table indicate the number of participants that meet the defined criteria. For example, 1786 participants tested positive for Giardia by ELISA were cases. 3469 participants who tested positive for Giardia by ELISA were controls. The marginal and grand totals are also calculated.

Contingency Tables
≡

☒ Plot Parameters
☐ Summary Statistics
☒ Contingency Tables
☐ Plot Grid
☐ Individual Plots
☐ Help

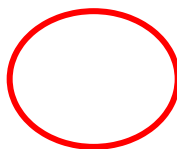
Contingency Tables

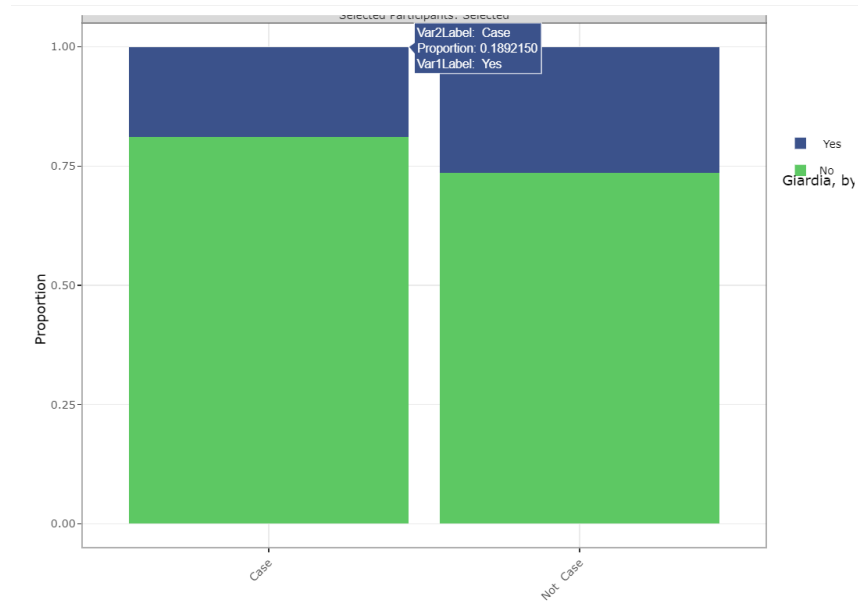
Facet(s): Selected Participants: Selected

	Case ↕	Not Case ↕	Totals ↕
Yes	1786	3469	5255
No	7653	9659	17312
Totals	9439	13128	22567

10. Clicking on 'Summary Statistics' shows the estimates Odds Ratios and Relative Risks and their p-values. Remember that these are just estimates and should not be over-interpreted. You'll notice that in these participants, testing positive for Giardia appears to be significantly associated with controls.

11. 'Plot Grid' will display a figure in green and blue which is a proportional physical representation of these counts. The blue box on the left hand side corresponds to the 1786 participants who tested positive for Giardia by ELISA and were cases while the green beneath it represents those that were Giardia negative cases. If you hover over any of the color shaded areas a pop-up box will tell you the attributes that you have specified and what proportion meet this group. In this case approximately about 19% of Cases are positive for Giardia.





8. Next take a look at another common pathogen. For Variable 1, select 'Cryptosporidium' from the drop down menu below. For Variable 2, keep Case.

Variable 1

Cryptosporidium, by ELISA

are / is

Yes

Variable 2

Case or Control Child

are / is

Case

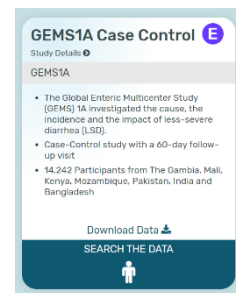
What is the association between Cryptosporidium and MSD? Is it significant? How does this differ from Giardia?

Summary Statistics			
Facet(s): Selected Participants: Selected			
	p-value	Odds Ratio	Relative Risk
Statistics	<0.0001	1.9679	1.8528
95% Confidence Interval	N/A	1.7921 - 2.1609	1.6873 - 2.0345

Finally, you may want to look at your associations by different strata, by site or by age group for example. To do this, click back to the 'Plot Parameters' tab.

Is there a significant association between *Cryptosporidium* and MSD in children less than 2 years of age? How about greater than 2 years of age? You can easily select different strata and define different groups.

9. If you have time see if you can replicate this analysis quickly in the GEMS1A dataset. Are the results similar? Again, the results of these simple contingency tables are not meant to be complete analyses but rather hypothesis generating tools. You may want to see if you see a similar pattern in other pathogens? The ability to point and click and change variables easily allows you to examine your data in ways that may never have occurred to you before!



You have completed the exercise for this section! Was this query of related Cases and Controls useful? What other concordant or discordant pairs might you want to look for? Please let us know if you have any questions or submit any comments via the 'Contact Us' link located in bottom right corner of any of our web pages or by emailing help@clinepidb.org