# Exercise 2. What is the average age of first asymptomatic parasitemia? (30 min)

In this exercise, we will use the PRISM cohort study to try and determine what the average age is the first time a child experiences asymptomatic parasitemia in Uganda. Whether someone infected with *Plasmodium falciparum* develops malaria is based on host genetics, parasite strain, and previous parasite exposure, which can lead to some immune system protection. It can therefore provide a clue as to the transmission intensity, with earlier ages of asymptomatic parasitemia associated with higher transmission rates. *If we want to determine what the first age of asymptomatic parasitemia is, what type of search do we need to do (Participant, Household, Observation, or Light trap)?*[1]

In this exercise you will walk through 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 learn how to revise a 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.

1. Navigate to http://clinepidb.org in your browser



---

[1] An Observation 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.

# Finding Filters and Examining Data in the Filters – Filtering for the first time a child presented with asymptomatic parasitemia

1. Start an Observations search by clicking the 'Observation' icon on either the study card or the drop-down menu under "Search a Study".
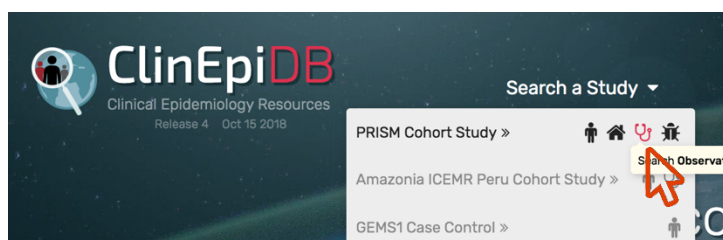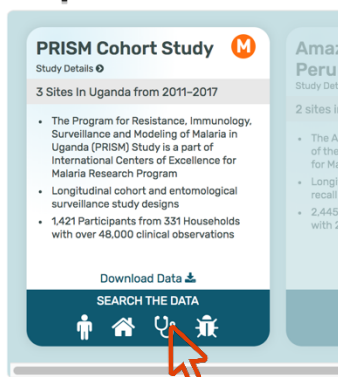


This will take you to a page with the Search Wizard at the top. The purpose of the Search Wizard is two-fold. First, it creates a simple way to categorize components of the data allowing for a step-wise approach to building searches. Second, it allows you to explore the data to see what the raw number and distribution of characteristics are in both the full dataset or filtered data.
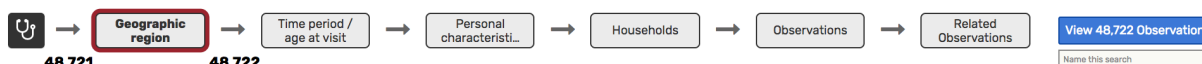


2. We want to filter the data for children. *Where can you find variables related to age?[2]*
3. Click on "Time period/age at visit" in the Search Wizard and select children under 12. Remember, you can select data by either entering the range into the indicated boxes or clicking and dragging your mouse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



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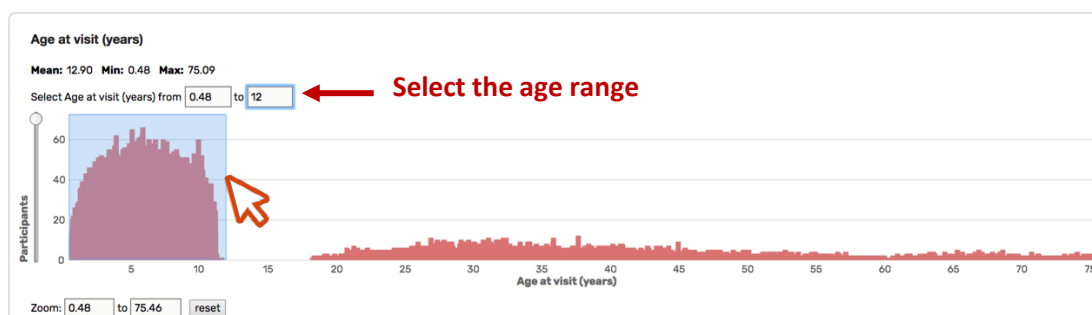
[2] Age-related variables actually appear in 3 places in the PRISM cohort study, under "Time period/age at visit", "Personal characteristics", and "Observations". In this case we will use the variable under "Time period/age at visit".

4. Move to the "Personal characteristics" step in the Search Wizard. It will open to "Date of birth" by default since that is the first filter listed on the left-hand side of the page. Note that all birthdates before 2000 are in grey, while all birthdates after 2000 are in pink. This is because we applied a filter. As we go through the Search Wizard, all data that meet previous filter criteria will be shown in pink, while data that do not meet the criteria will be shown in grey. Since we filtered on children, only participants born in 2001 or later are selected by our filter.
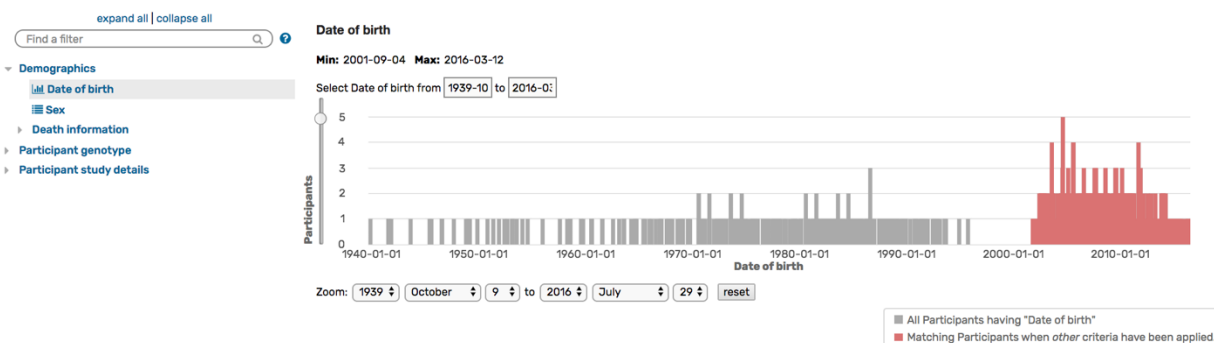


Note: You can always click on an earlier step in the Search Wizard and add or remove filters. However, filters are only applied in the order of the Search Wizard, so if you filter based on a variable under "Observations", you will not see the data under "Personal characteristics" update to reflect that.

5. Click through the available filter categories on the left-hand side of the page if you haven't already. Clicking a category will reveal the filters within. Clicking the 'expand all' link at the top of the list will reveal all the filters in each category at once.

**Click to see all filters in the step**

**Search for a specific filter by typing in this field**
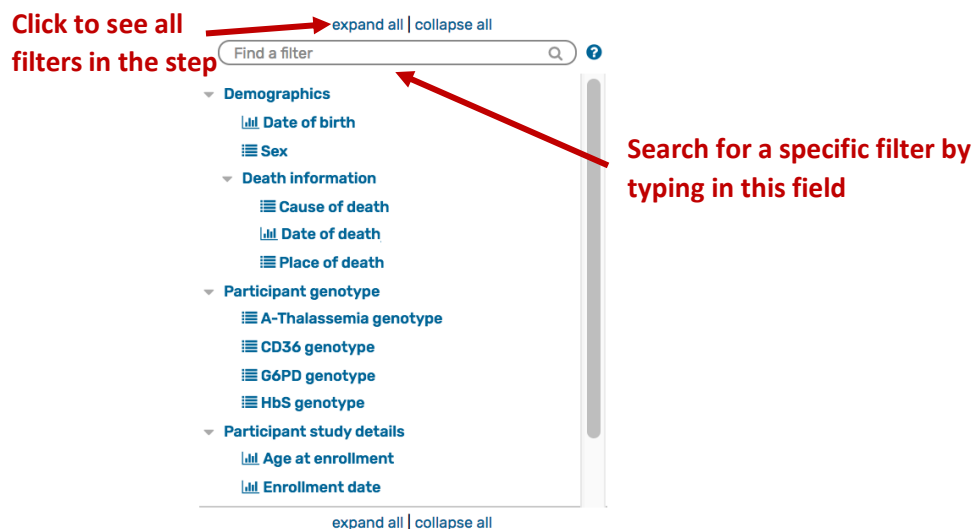


6. Now let's find malaria filters. Click on "Observation" in the Search Wizard and look through the available filters. For this example, we want to use "Malaria diagnosis and parasite status". You can find a specific filter by typing its name into the "Find a filter" search box.

7. We are interested in identifying all observations where a child was infected with *Plasmodium* but did not show symptoms. *Which values should we select?*[3]



**Mouse over to learn more**

8. Notice that the numbers in the second column (Remaining Observations) do not match the numbers in the third column (Observations). *Why is this? (Hint: Learn more by mousing over the question mark)*[4]
9. 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.*[5]
10. 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.



Once you enable the filter, you will see all the filters from the "Observations" step. The key difference between the two steps is that this filter works by allowing you to keep or remove observations you already filtered for based on a second set of observations that you define in this step. In this case, we are trying to identify the first observation of asymptomatic parasitemia. Right now we have all observations for children that were ever found to have

---

[3] Based on the data, we are interested in children who were either "Blood smear negative/LAMP positive", which indicates sub-microscopic infections, or "Blood smear positive/no malaria".

[4] Remaining observations reflects only the data that match all the filters you've applied, while Observations reflects all the data.

[5] 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.

asymptomatic parasitemia, so we want to remove all observations of asymptomatic parasitemia except for the first one. How do you think we can do this using the "Related Observations" step?

11. 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 did our "Observation". Click on the filter for "Malaria diagnosis and parasite status" and select the values of "Blood smear negative/LAMP positive" and "Blood smear positive/no malaria".

12. 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 ×

**Malaria diagnosis and parasite status**

⬤ Keep checked values at top                                                                 **1 of 48,722** Observations have no data provided for this filter

| | Malaria diagnosis and parasite status | Remaining Observatio... ❓ 48,721 (100%) | Observatio... ❓ 48,721 (100%) | Distribution ❓ | % ❓ |
|---|---|---|---|---|---|
| ☐ | 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%) |

Filter tree (left panel):
expand all | collapse all
Find a filter 🔍 ❓
▼ Observation
  ≣ Malaria diagnosis and parasite status
  ≣ Visit type
▸ Diagnosis
▸ Hospitalization
▸ Observation details
▸ Personal vector intervention
▸ Physical examination
▸ Signs and symptoms
▸ Treatment
▸ Sample

13. Now we are left with 854 observations of the first asymptomatic parasitemia in children. Click on the blue button in the Search Wizard to visit the results page and learn more.

## Analyzing Data via the Results Page – What's the average age of first asymptomatic parasitemia?

1. From the results page you can add and remove columns, sort columns, and view histograms of the filtered data.

**854 Observations** [Revise] [Hide search strategy panel] [Save] [Share]

Sort    Remove column    View histogram    **Add columns**

[Observation Results] [Analyze Results]

First 1 2 3 4 5 Next Last   [Advanced Paging]                     Download  Add to Basket  [Add Columns]

| Observation Id | Visit type | Date of visit | 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 | NA | NA |
| 100619044 | Scheduled visit | 2012-02-25 00:00:00.0 | 8.68 | Blood smear negative / LAMP positive | NA | NA |

2. Bring up the histogram for "Age at visit (years)". *What is the average age?[6]*

---

[6] The average age at first asymptomatic infection is 4.956

3. 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? (Hint: hover over the bars to see extra information)[7]*
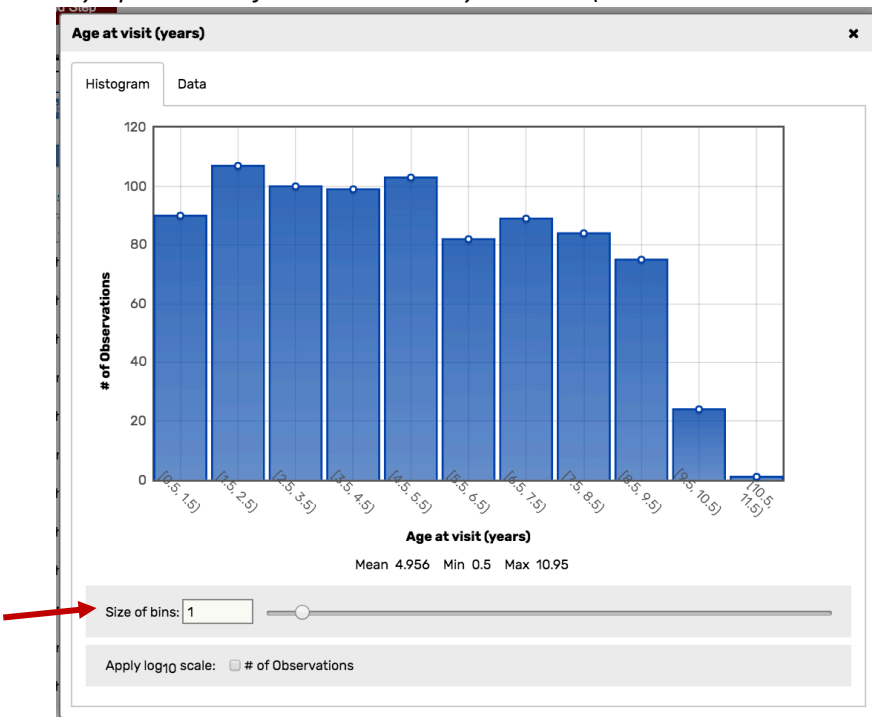


4. We know that the three different PRISM sites were selected to represent different transmission settings. *How would you expect the average age at first asymptomatic visit in Nagongera, a high transmission setting, to compare to Walukuba, a lower transmission setting?[8]*

## Revise a Search Strategy – How does the average age of first asymptomatic parasitemia vary across sites?

Instead of starting a whole new search from scratch, it's possible to edit a search strategy by clicking the "Revise" button from the results page.



---

[7] 107 children

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

In this example, we will 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. Click on "Duplicate" on the right-hand side. You will see a pop-up asking if you want to make a copy of the strategy. Click "OK".

3. Now you should see two searches in the search strategy panel. You can rename the new search by clicking on its name or by clicking "Rename". Rename the strategy to "Nagongera".

4. 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.
5. Select "Nagongera" for the filter "Sub-county in Uganda". 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?*[9]



6. Now click the blue "View Observations" button to return to the results page. *What is the average age at visit now? How does this compare to the average of 4.956 that we saw when looking at data from all 3 sites? Is this what we predicted?*[10]
7. We would also like to see what the average age is in Kihihi and Walukuba. Revise your search strategy and figure out the average age at first asymptomatic observation for these two sites. *What is the average age at visit for Kihihi? For Walukuba?*[11]

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? How could we adapt the search to correct for this?*[12]

A second thing to consider is some of these observations may be infections that were caught before they became symptomatic. To remove infections that became symptomatic, we will need to use a multi-step strategy. Check out the bonus section below on how to do this if you have time.

---

[9] 349 observations of first asymptomatic infections
[10] 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.
[11] 5.453 years in Kihihi, 4.956 years in Walukuba
[12] 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. 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.

## Bonus Section: Instituting a Multi-step Search Strategy – Removing observations from children who became symptomatic (15 min)

We want to adapt our original search to remove observations where the subject converted from asymptomatic to symptomatic infection shortly after (within 2 weeks). This requires adding a separate step where we identify all asymptomatic infection observations that became symptomatic in 2 weeks, then subtracting those results from the results of our original search.

1. Start by clicking the "Add Step" box for the original search which returned 854 observations.
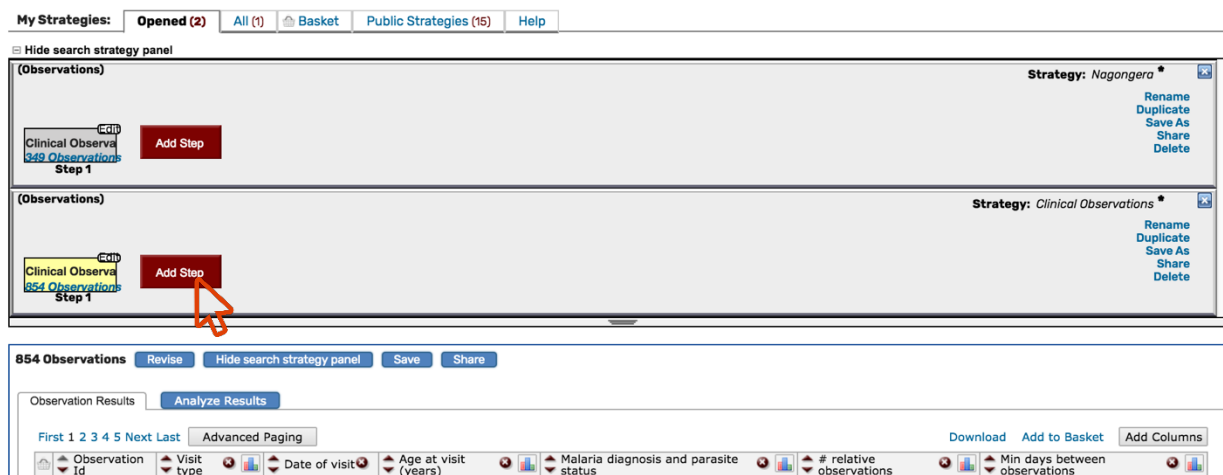


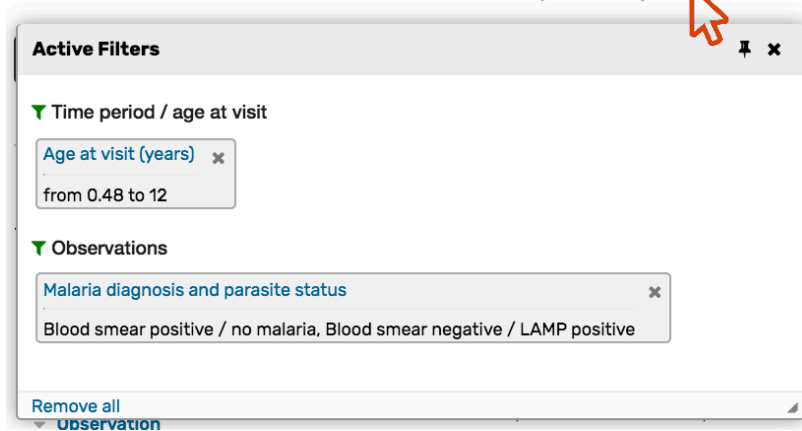2. Choose "Select a set of clinical observations (PRISM)".
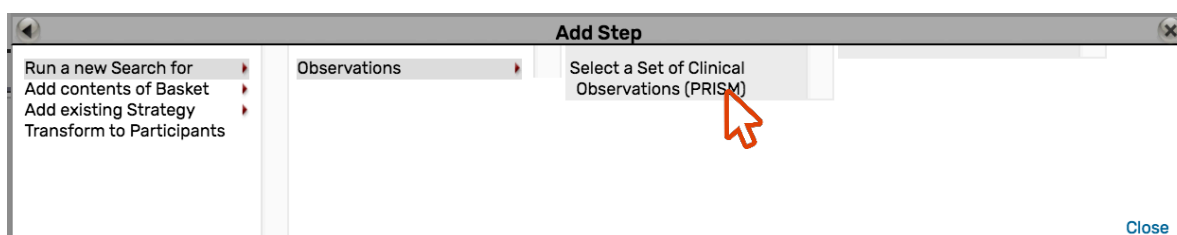


3. Notice that the panel that opens up looks exactly the same as the Search Wizard you used earlier, except for the new section at the top. This section allows you to choose how to combine this search (search 2) with the original search (search 1). In this case we want to identify all observations that became symptomatic and remove them. *Which operation do you think we should use?[13]*



---
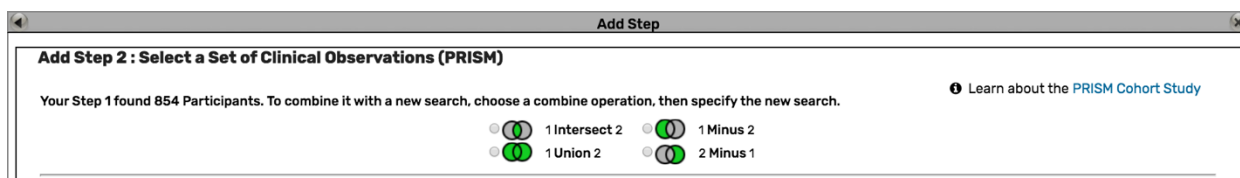
[13] 1 Minus 2

4. The first part of this search should be similar to what we did before. We want to identify children with asymptomatic infection. Under the step "Time period/age at visit" select children aged 0.48 to 12 years old. Then under "Observations" click on "Malaria diagnosis and parasite status" and select 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:



5. Now go to "Related observations". We want to keep all observations that become symptomatic. *Which value should you select for "Malaria diagnosis and parasite status"?[14]*
6. We want to keep observations that became symptomatic up to 14 days after our original observation. *How can we specify this under "Related observations"?[15]*
7. There should be 634 observations. Go to the results page by clicking the blue "Combine 634 observations" button at the end of the Search Wizard. *How many observations of first asymptomatic infection are there now that you've removed asymptomatic infections that converted to symptomatic infections?[16]*
8. 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.[17]*

How did the average age change compared to our first search? Did you expect this? Why or why not?

You have completed the exercise for this section! Great job! Please let us know if you have any questions or comments via the 'Contact Us' link located in the bottom right corner of any of our web pages. You can also email us at help@clinepidb.org.

---

[14] Choose "Symptomatic malaria"
[15] **Keep** observations that are **1** to **14** days **before** the related observations specified below
[16] 745 observations
[17] Average age is 5.043 years