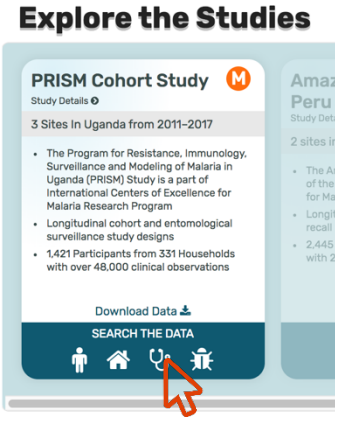
How does house design relate to malaria incidence in children in Uganda? (25 min)

In this exercise we will use the PRISM cohort study to determine if there is an association between house design and risk of *Plasmodium* infection. The PRISM study collected variables such as wall, floor, and roof type, the type of eaves used, and the presence of airbricks. You will explore the data using the observation search. You will also use the Analysis Apps to visualize the data while trying to answer the question of how house design relates to malaria in children in Uganda.

Start an Observations search

We want to determine how house design relates to malaria incidence in the PRISM cohort study. This study collected data from all participants every three months and from a subset of children every month. Participants were also encouraged to come to the clinic any time they were sick, so each participant has multiple observations. *What type of search do you think we want to do based on this information? Why?*[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.



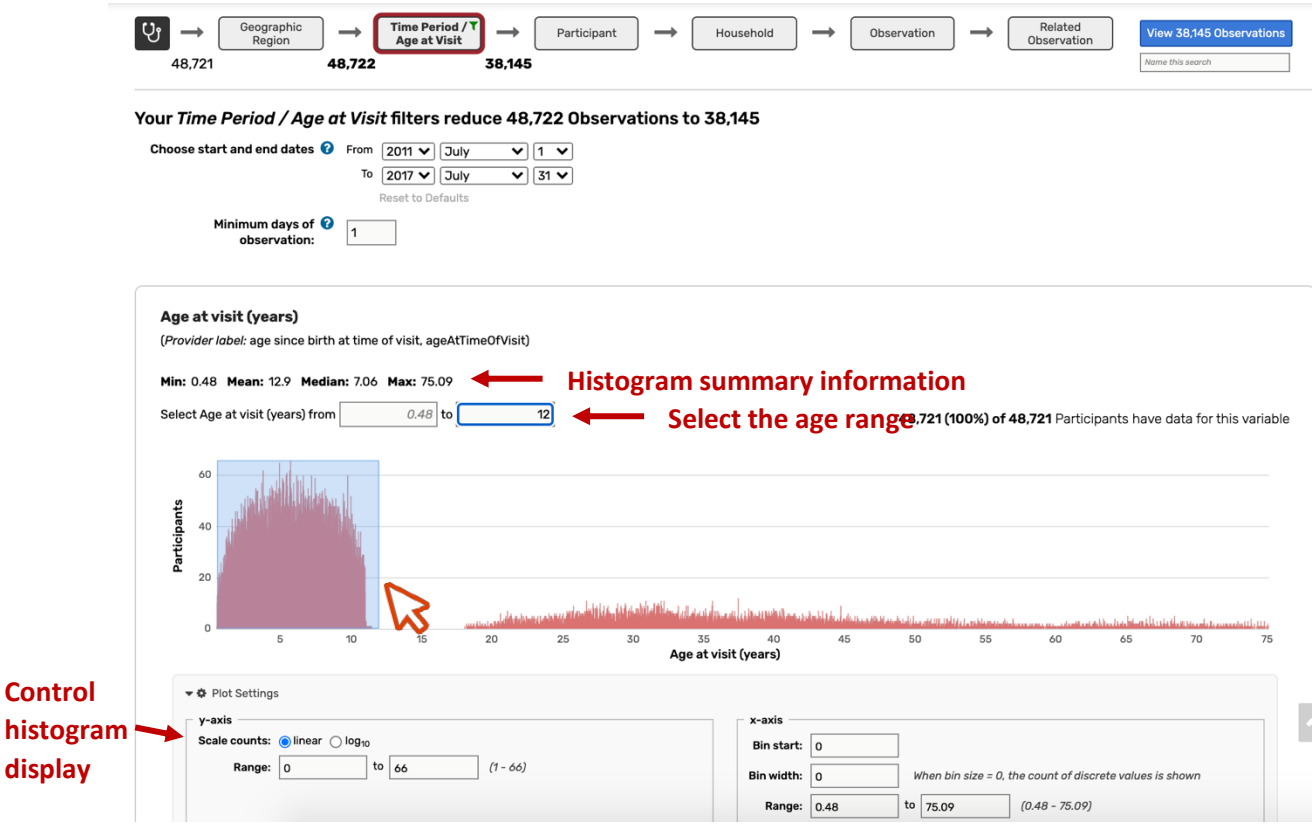
Finding and Examining Data Variables – Determining what house design and malaria-related variables were collected

[1] An Observation search will return one row of data per observation and allow us to capture data from every individual malaria episode.

We are interested in looking at how house design relates to malaria in children. *Where might you expect to find age-related variables in the Search Wizard? How about house-related variables? And malaria variables?*[2]

1. Click on "Time period/age at visit" in the Search Wizard. *Can you find the maximum Age at visit (years)? How about the average?* [3]
   The data for this variable has a very interesting distribution pattern. The Study Page notes that the study enrolled all children from age 6 months to 10 years and one primary caregiver per household.

2. Select the children you want to subset the data (<=12 years old) on by either typing the age range you want into the boxes or by clicking and dragging on the graph to select the range of interest with your mouse.



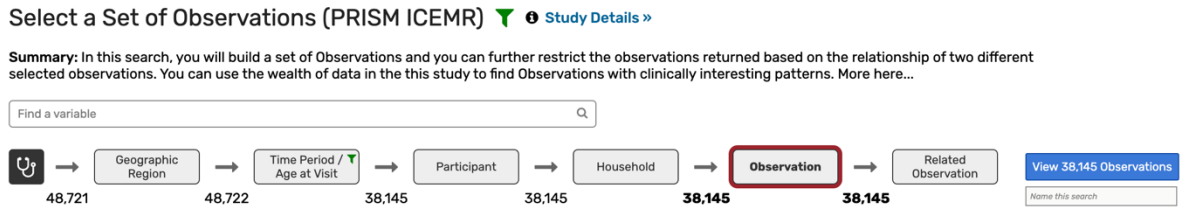
3. Next, take a look at variables that relate to house design by clicking on "Households" in the search wizard. Examine the available categories in blue text on the left-hand side of the page. Clicking a category reveals the variables within. Clicking the 'expand all' link at the top of the list will reveal all

[2] Time period/age at visit and Participant nodes have variables related to age, the Household node has variables related to house design, and the Observation node has variables related to malaria episodes

[3] The max age at visit is 75.09, the average is 12.90.

the variables in each of the subcategories. Scroll down and read through them. *How many households have open eaves? How many households have mud walls?*[4]

4. Now let's look at malaria-related variables. Use the "Find a variable" search and look for "malaria". Click on "Malaria diagnosis and parasite status". *What type of data does it contain?*[5]



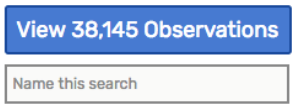
Using the Analysis Apps to Visualize Data

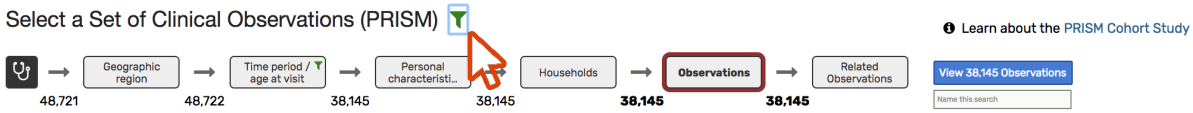
We have 3 different Analysis Apps that are available for visualizing data – Data distributions, Contingency tables, and Data summaries. Each app graphs the data in a different way. We want to know how house design is associated with malaria, so let's start by using Data distributions to look at how malaria diagnosis varies based on the house characteristic of "airbricks" and by site.

An important warning about these applications; they are intended for data exploration only! You may choose to examine any number of variables or relationships between variables using these apps but all results should be considered hypothesis generating. More complete, rigorous statistical analyses with proper consideration of study design, possible sources of bias and confounders is strongly recommended before any analyses are considered complete.

Data Distributions – How does malaria diagnosis in children vary based on the type of airbricks used?

1. Go to the results page by clicking the blue button at the far right-hand side of the wizard. Note: if your now button says something other than "View 38,145 Observations", you might have applied additional selections. To see all the variables that you have subsetted on, click on the green filter icon at the top of the Search Wizard then remove any extra filters.

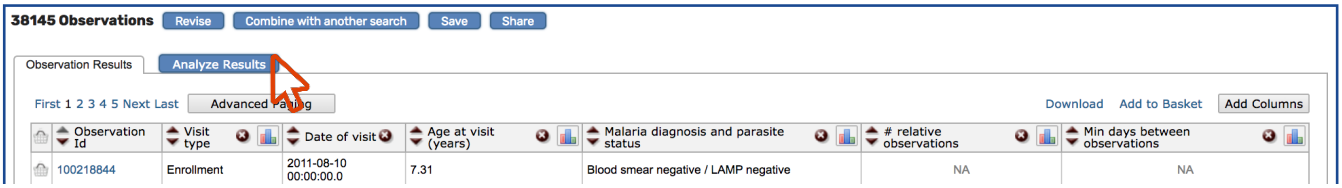




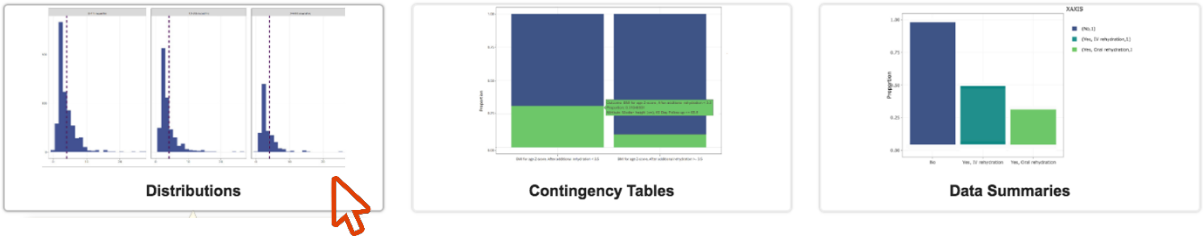
[4] 116 households have open eaves, 202 households have mud walls

[5] Values indicate whether the participant had a positive or negative blood smear or LAMP, diagnosing *Plasmodium* infection. Symptomatic malaria means they had symptoms + a positive blood smear

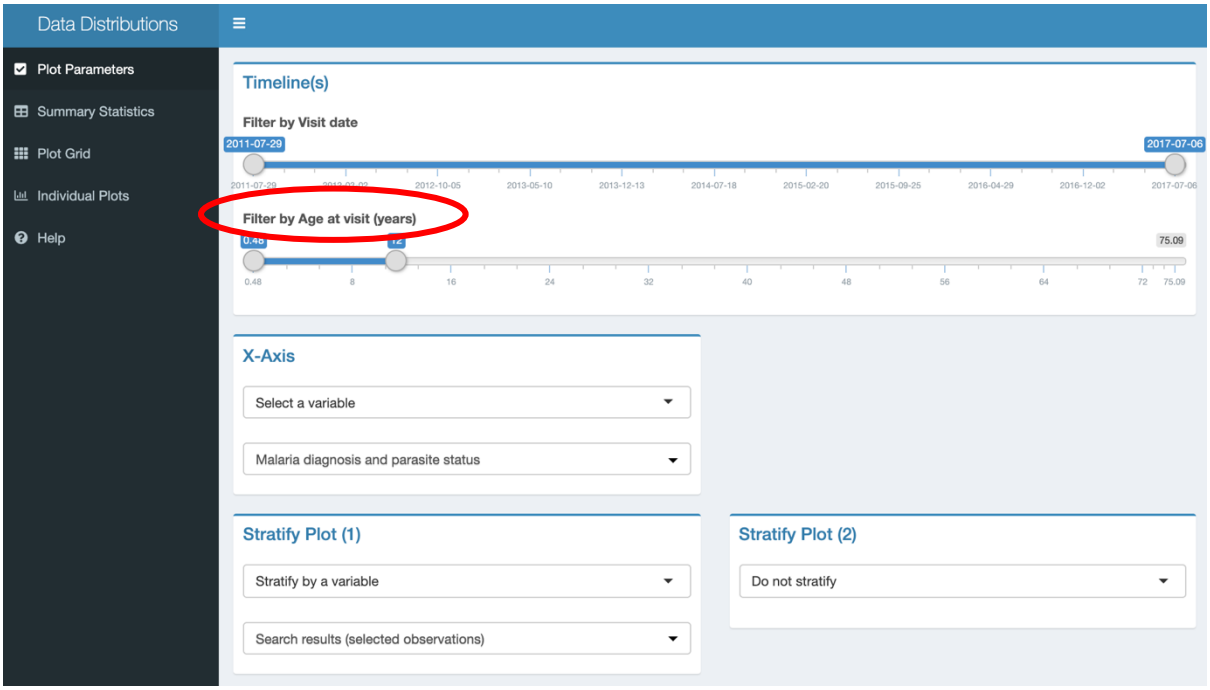
2. Click on the “Analyze Results” button from the results page



3. Select the “Distributions” analysis tool

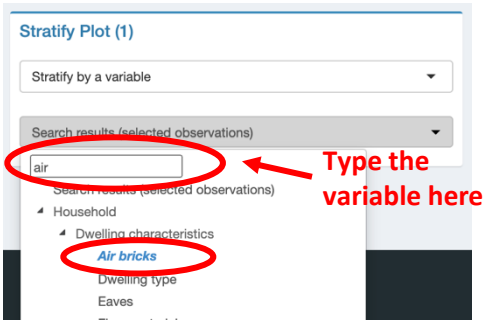


4. By default, the Analysis Apps work on all the data, not just the data you selected in the Search Wizard. However, you can choose to stratify your data based on your selected results. Here, we are interested in looking at how the Malaria diagnosis and parasite status varies for children based on the presence of air bricks in their home and which site they are from, which will require stratifying by 2 different variables. Instead of using one stratification parameter for choosing only the selected results, we will use the Filter by Age at visit (years) parameter to select for children under 12.



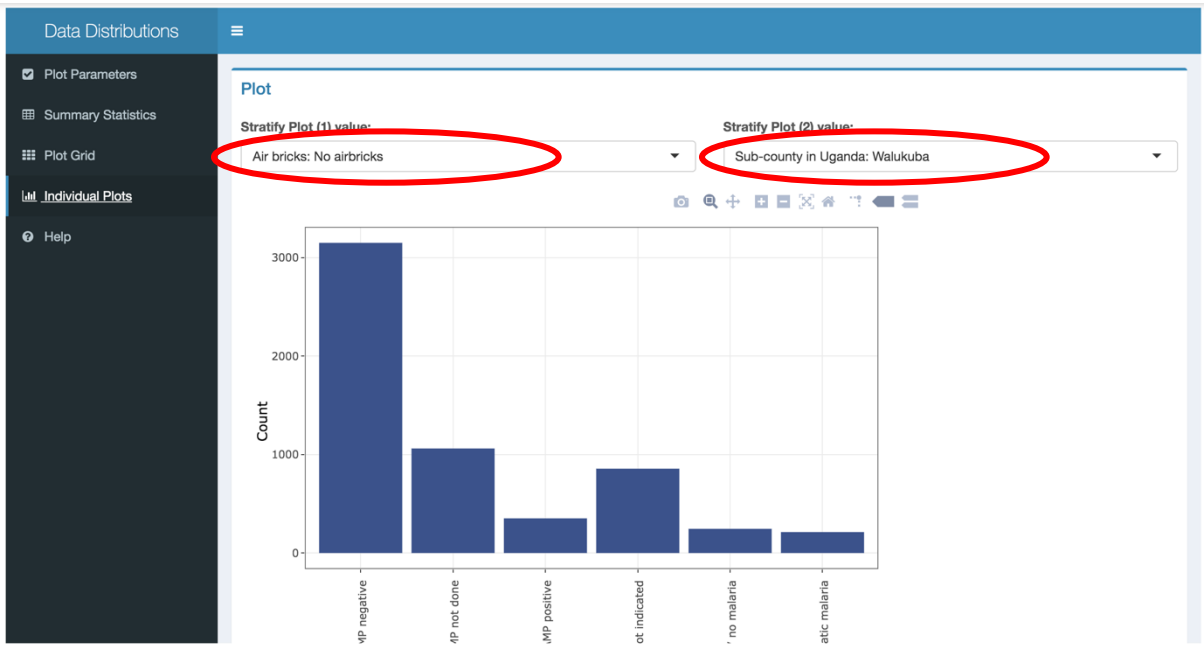
5. In this analysis, we want to see each observation of malaria diagnosis and parasite status tallied, so we will keep the X-Axis as “Malaria diagnosis and parasite status”.

6. To see how “Malaria diagnosis and parasite status” varies based on “Air bricks”, set up Stratify Plot (1) by changing the first field to “Stratify by a variable” and the second to “Air bricks”. You can either click through the categories to find your variable or start typing it in the box at the top.



7. The PRISM study chose 3 sites with different transmission intensities. A relationship between house design and geographic region might therefore confound results, so let’s set up the second strata for “Sub-county in Uganda”.

8. Take a look at Individual Plots and use the drop-down menus at the top to select which distribution plot you want to look at. Hover over the bars to see the specific count. *How many observations of symptomatic malaria are associated with no airbricks in Nagongera? With unscreened airbricks?*[6] *What can you say about the use of airbricks across the different sites?*



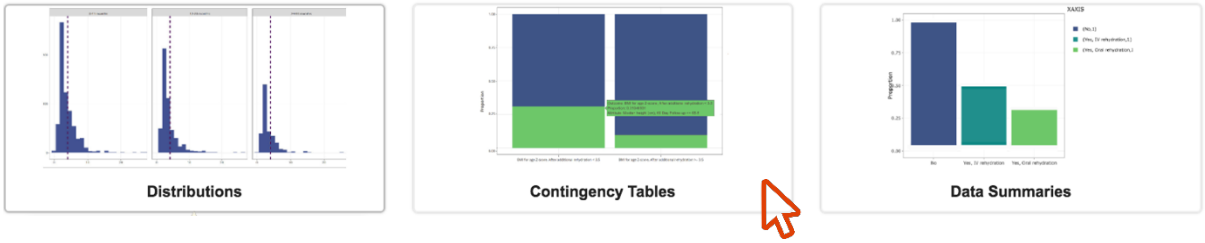
By looking at the Plot Grid, we can see that the most common type of airbrick varies by site. This Analysis App lets us see how the data breaks down by different variables. Take a moment to try stratifying by other variables that you are interested in. Next, we’ll use the Contingency Table app to generate 2 x 2 contingency tables and calculate some basic statistics.

[6] 2,159 observations of symptomatic malaria with no airbricks in Nagongera, 759 observations of symptomatic malaria with unscreened airbricks in Nagongera

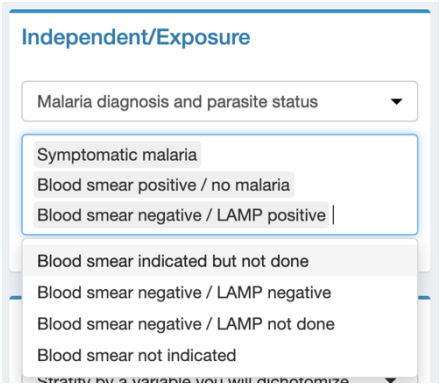
Contingency Tables – What's the risk of infection for a child living in a modern vs. traditional dwelling?

The contingency tables app allows you to look at the association between two variables and calculates a p-value, odds ratio (OR), and risk ratio (RR) with 95% confidence intervals.

1. Click on the "Analyze Results" button and select the Contingency Tables app.



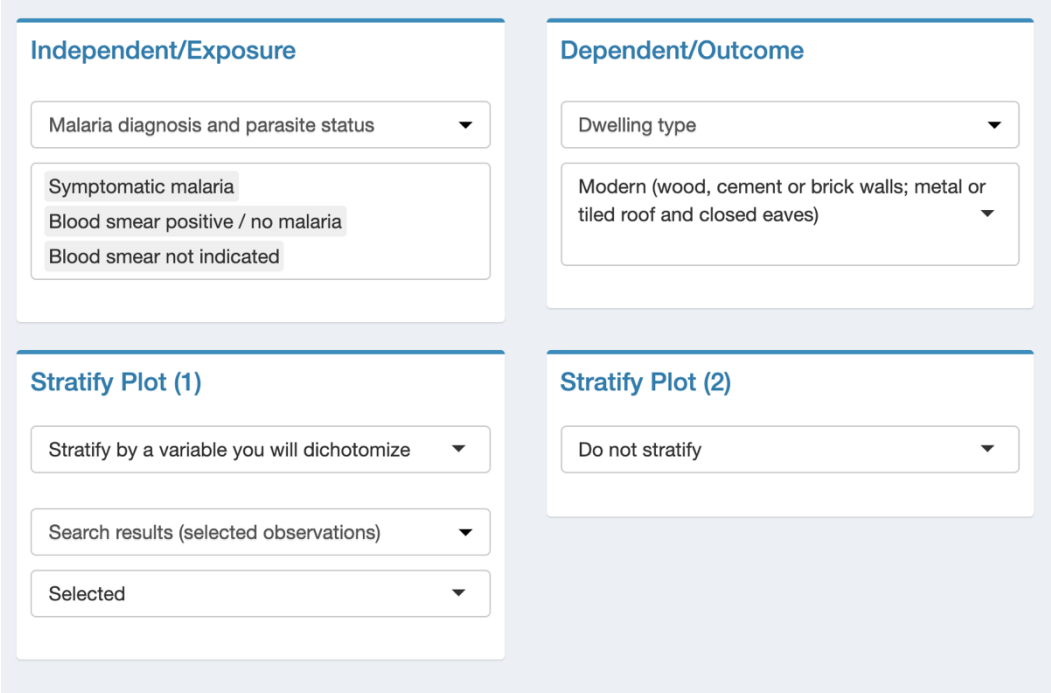
2. We want to look at the relative risk of infection for a child living in modern vs. traditional dwellings and use the variable "Malaria diagnosis and parasite status". However, a 2x2 contingency table only allows two values per variable, and there are more than two potential values for "Malaria diagnosis and parasite status". Can we still use this variable?



   Yes, we can still use the variable, but we have to group the values. Here, we are interested in all observations where a child was found to be infected compared to all observations where the child was not found to be infected. We will therefore select "Symptomatic malaria", "Blood smear positive/no malaria", and "Blood smear negative/LAMP positive" from the drop-down menu. The remaining values ("Blood smear indicated but not done", "Blood smear negative/LAMP negative", "Blood smear negative/LAMP not done", and "Blood smear not indication") will automatically become the second group.

3. The PRISM cohort study derived the variable of "Dwelling type" based on the wall and roof material and eaves. Let's select this as our second variable and look at modern homes.

4. Stratify Plot (1) is set to stratify based on "Search results (selected observations)" by default. *If we didn't stratify the data this way, how might we be sure to only use the same data as what we selected in the Search Wizard? Why won't this always work?*[7]



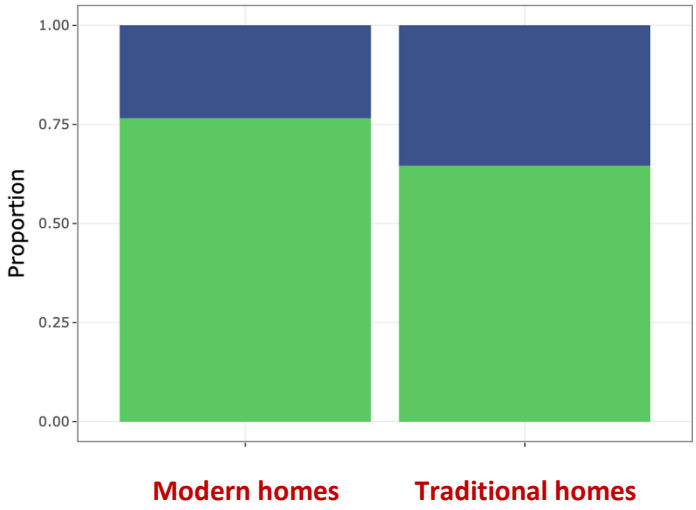
5. Look at the "Summary Statistics". *Is a modern home protective against Plasmodium infection? What's the odds ratio and confidence interval for a child getting infected if they live in a modern home compared to a traditional home?*[8]

6. Check out the plots as well. In this plot of selected (the children you filtered on), blue shows observations of infection, while green shows observations without infection. You can see there are significantly fewer infections (blue) for children living in modern homes compared to traditional homes.

7. What other variables are you interested in? Take a look at some of them.

[7] Since we only selected for children in the Search Wizard, we could use the "Filter by Age at visit (years)" parameter at the top of the Contingency Tables app. However, if we had applied multiple selection criteria, we would have to stratify based on "Search results (selected observations)" in order to restrict our analysis to just the data of interest.

[8] A modern home is protective against *Plasmodium* infection. The odds ratio is 0.4393, with a 95% confidence interval from 0.4152-0.4648.



Do you think house design alters children's risk of *Plasmodium* infection? What other variables would you want to control for? Are there any intervention recommendations you would make based on this data? 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.