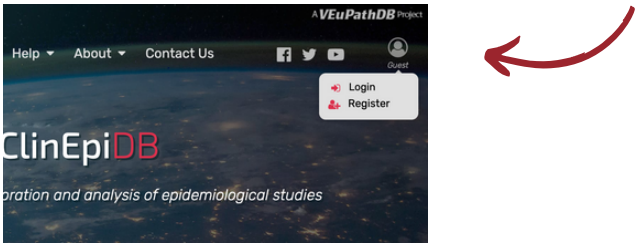
Exploratory data analysis on the ClinEpiDB Beta Platform
A structured exercise

1. **Saving your work**: If you would like to save your analysis, make an account by clicking the "Register" link under the "Guest" icon (top right). It only takes a minute to register!



You may also continue as a guest without making an account; the analysis will be accessible as long as the browser window is open.

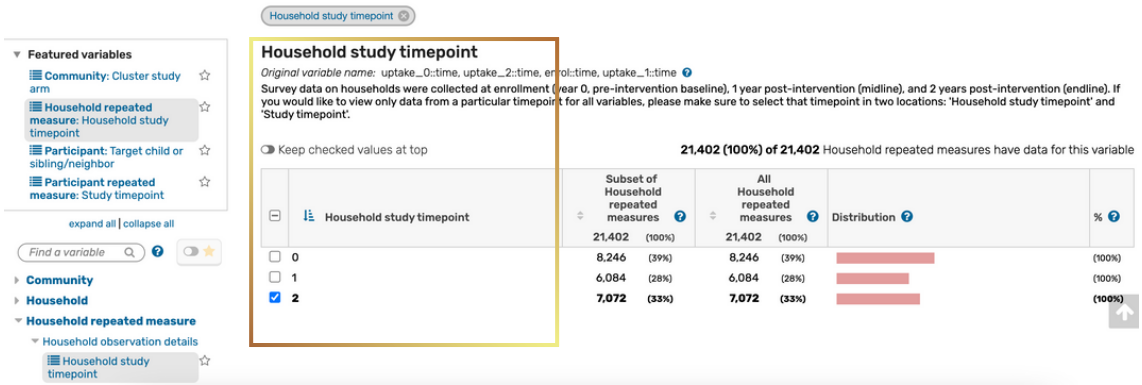
2. Go to the **WASH benefits Kenya** study page. Click the "**View study details**" tab and read the summary. In this trial, clusters of households were randomized to one of several WASH (water, sanitation and hygiene) intervention arms or to control arms. One of the primary outcomes evaluated was diarrhea in target children.

Hypothesis: We hypothesize that target children belonging to intervention arms have reduced diarrhea as compared to target children in control arms. Interventions can succeed only if there is sufficient uptake in the target population, so we may also want to evaluate the uptake of the interventions.

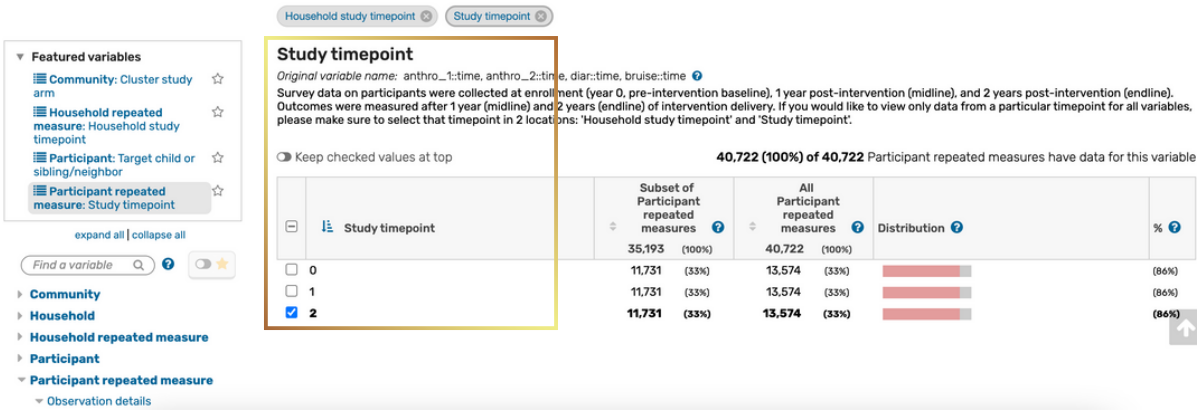
3. Choose an appropriate subset of data for this exploratory analysis

a. Click the "**Browse and subset**" tab

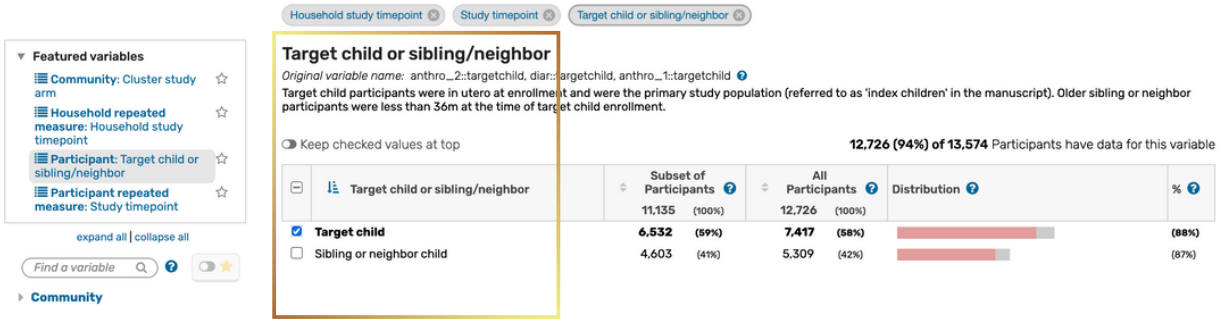
b. We are interested in the outcome of this trial, so how can we restrict our subset to the endpoint in the two timepoint variables, at the household and participant level?



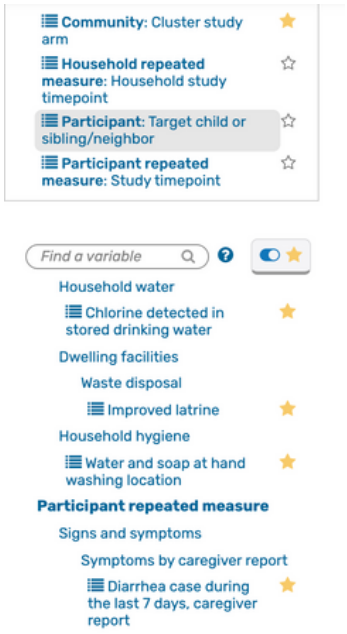
Exploratory data analysis on the ClinEpiDB Beta Platform
A structured exercise (continued...)



c. We want to evaluate diarrhea specifically in target children. How could we choose this subset of participants?



4. Identify variables of interest for this analysis



a. **Intervention** - Which variable tells you what intervention clusters were randomized to receive? Star that variable.

b. **Intervention uptake** - Which variables under "Household repeated measures" in the variable tree assess uptake of WASH interventions? Star a few of these.

c. **Outcome** - What variable should we use to assess the outcome of diarrhea? Star the variable of interest.

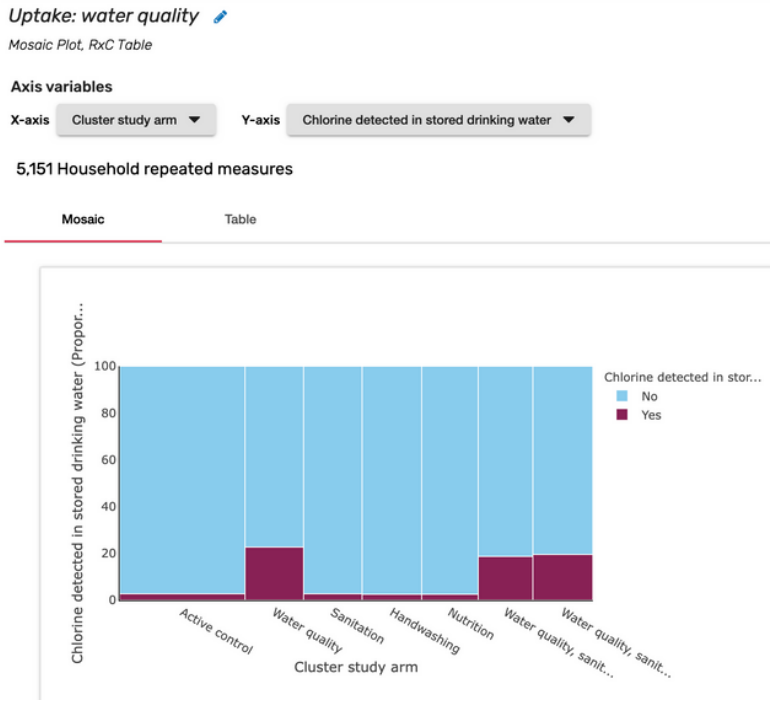
You may end up with a list of starred variables that looks like the one on the left.

Exploratory data analysis on the ClinEpiDB Beta Platform
A structured exercise (continued...)

5. Create visualizations of the data

a. Click on the "**Visualize**" tab to plot data

b. We want to plot the uptake of a WASH intervention identified above against the cluster study arm. Are these continuous or categorical variables? What visualization tool should we use?



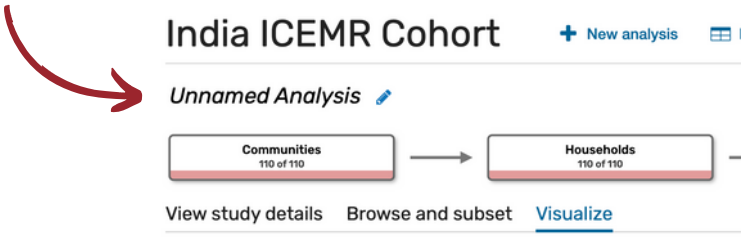
c. Similar to the example above, make the following plots using the Mosaic plot RxC tool and name each one:

i. X axis: Cluster study arm; Y axis: Chlorine detected in stored drinking water
ii. X axis: Cluster study arm; Y axis: Improved latrine
iii. X axis: Cluster study arm; Y axis: Water and soap at hand washing location
iv. X axis: Cluster study arm; Y axis: Diarrhea case during the last 7 days, caregiver report

What do each of these plots say about the uptake of the intervention across study arms? Does the diarrhea prevalence differ across study arms based on your plot?

Indeed, the published study results show that none of the interventions reduced diarrhea prevalence compared with the active control.

6. Name your analysis.



If you check the dropdown menu Workspace > My analyses, you will see that the analysis automatically appears in this table.

Thank you for completing this exercise on performing an exploratory data analysis on beta.clinepidb.org!