

## Exercise: How does poverty impact the use of bednets in Rourkela and Chennai? (40 min)

In this exercise we will use the India ICEMR longitudinal cohort study to determine if there is an association between poverty and bednet usage. You will learn how to access the data and explore it using the observation search filters. You will also discover how to use the Shiny Apps to visualize the data while trying to answer the question of whether poverty is associated with bednet usage.

Navigate to <http://clinepidb.org> in your browser:

The screenshot shows the ClinEpiDB homepage. At the top, there's a navigation bar with links for 'Search a Study', 'Workspace', 'Community', 'Contact Us', and social media icons. Below the header is a banner with the text 'Welcome To ClinEpiDB' and 'Advancing global public health by facilitating the exploration and analysis of epidemiological studies'. The main content area is titled 'Explore the Studies' and features three study cards: 'GEMS1 Case Control', 'GEMS1A Case Control', and 'India ICEMR Longitudinal Cohort Study'. The 'India ICEMR Longitudinal Cohort Study' card is highlighted with a red box around its 'SEARCH THE DATA' button. To the right, there's a 'News' sidebar with recent announcements like 'ClinEpiDB 5 Released' and 'ClinEpiDB at ASTMH'.

### Start a Participant search

Take a look at the India ICEMR Cohort Study card (boxed in red above). Notice the 3 icons located under the words 'Search the Data'. Each icon starts a different type of search, Participant, Observation, or Household. Within each search, you can filter the data based on any of the variables that were collected in the study, the only difference is in the data tables you get in the end. A Participant search will return one row of data per study participant, an Observation search will return one row of data per observation, etc.

We want to determine how poverty relates to bednet usage in the India ICEMR cohort study. This study collected data from all participants every three months, but questions about bednet usage were only

asked once, at the beginning of the study. *What type of search do you think we want to do based on this information? Why?*<sup>1</sup>

## Finding Filters and Examining Data in the Filters – Determining what variables were collected

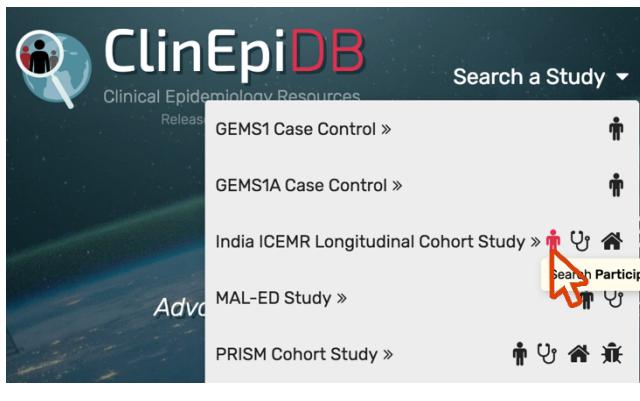
1. Start a Participant search by clicking the ‘Participant’ icon on either the study card or the drop-down menu under “Search a Study”. This will take you to a page with the Search Wizard at the top.

**Study card**



A screenshot of the ClinEpiDB study card for the India ICEMR Longitudinal Cohort Study. It shows a participant icon with a red box around it, indicating it's the target for a click. Other icons include a download data icon, a search the data button, and a home icon.

**Drop-down menu**



A screenshot of the ClinEpiDB homepage showing the 'Search a Study' dropdown menu. A participant icon with a red box around it is highlighted, indicating it's the target for a click. Other options in the menu include GEMS1 Case Control, GEMS1A Case Control, India ICEMR Longitudinal Cohort Study, MAL-ED Study, and PRISM Cohort Study.

2. 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.

Select a Set of Participants (India ICEMR)

[Learn about the India ICEMR Longitudinal Cohort Study](#)



The search wizard interface for the India ICEMR study. It shows a sequence of steps: 'Geographic region' (highlighted with a red box), 'Households', 'Personal characteristi...', 'Observations', and 'Related observations'. Below each step is a small icon. The 'Observations' step has a black square icon with the number '397' below it. To the right of the steps is a blue button labeled 'View 397 Participants' and a text input field 'Name this search'.

Notice that there is a number below the black square Observation icon. This number represents the total number of participants that are included in this dataset.

3. We are interested in looking at how poverty relates to bednet usage. Spend some time looking at what variables were collected. To do this, click on the different grey buttons in the search wizard, then click “expand all” to see all filters within that step. You can also start typing into the

<sup>1</sup> A Participant search will return one row of data per participant and allow us to capture data from every participant on bednet usage.

"Find a filter" box. Try typing "poverty line" and click on the result. *What percent of observations are recorded as "Above" the poverty line?*<sup>2</sup>

### Select a Set of Participants (India ICEMR)

[Learn about the India ICEMR Longitudinal Cohort Study](#)

No Households filters applied yet

**Click to see all filters in the step** [expand all](#) [collapse all](#)

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

**Household data collection date**

Min: 2012-03-18 Max: 2014-05-14

Select Household data collection date from  to

Zoom:    to

All Observations having "Household data collection date"  
Matching Observations when other criteria have been applied.

4. Keep looking through the variables available under different steps. Search for "bednet" under the Household step. Then search for bednet under the Observations step. *How are these variables different?*<sup>3</sup>
5. Go back to the Households step and find the "Poverty line" variable. Select "Below".

[expand all](#) [collapse all](#)

**Poverty line**

Keep checked values at top

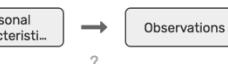
110 of 220 Observations have no data provided for this filter

	Poverty line	Remaining Observatio...	Observatio...	Distribution	%
<input type="checkbox"/> Above	54 (49%)	54 (49%)	54 (49%)	(100%)	
<input checked="" type="checkbox"/> Below	56 (51%)	56 (51%)	56 (51%)	(100%)	

<sup>2</sup> 49% of observations are above the poverty line.

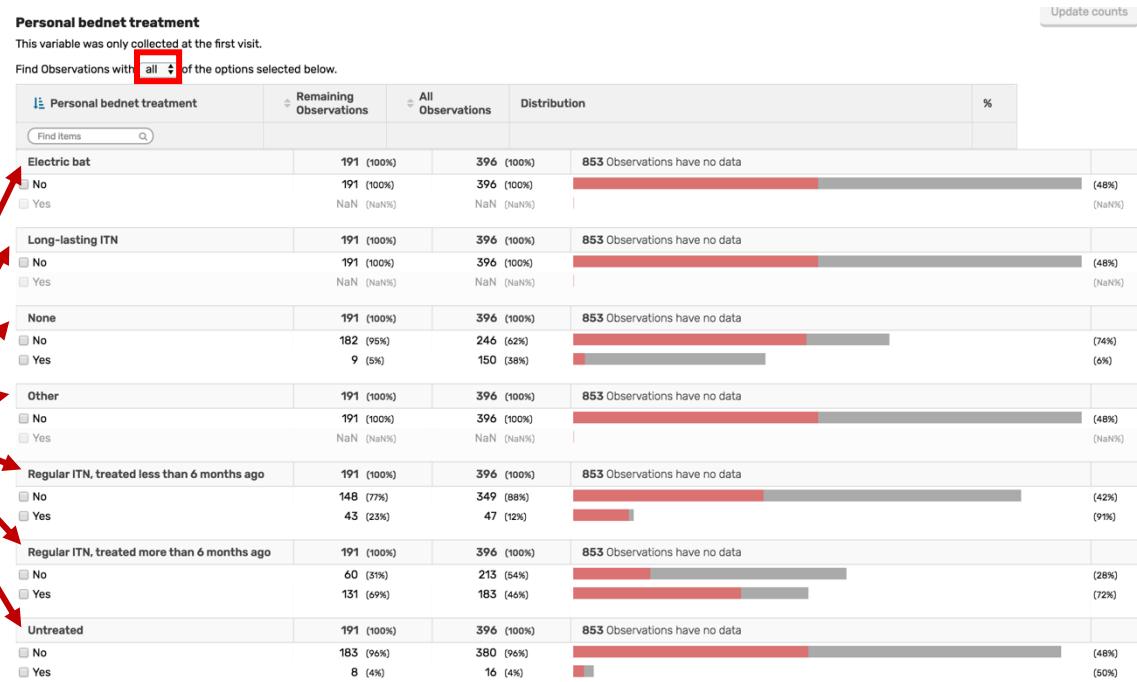
<sup>3</sup> The variables under Household collect the number of each type of bednet that belong to a household. The variables under Observations collect what type of bednet an individual participant reported using at the first visit.

6. Notice that the numbers in the Search Wizard have updated based on the filter we applied. *How many participants are now selected? Why doesn't the number match what you see on the poverty line table?*<sup>4</sup>

Select a Set of Participants (India ICEMR)     

1 Learn about the India ICEMR Longitudinal Cohort Study  
 397 → **Households**  191  
 ? Refresh counts   
 Name this search

7. Now let's look at the impact of a household being below the poverty line on which type of bednet a participant uses. *Which variable do we want to look at and what step is it under?*<sup>5</sup>
8. Go to the Observations step then find the variable “Personal bednet treatment”. This type of variable is called a “multifilter”. It includes multiple variables (Electric bat, Long-lasting ITN, None, etc.) that all appear on one screen. The parameter box at the top allows you to choose whether you want to find participants with “ALL” the values you select below or “ANY” of the values you select below. *How could you identify participants who reported that they did not use an untreated net or a regular ITN that was treated more than 6 months ago?*<sup>6</sup>

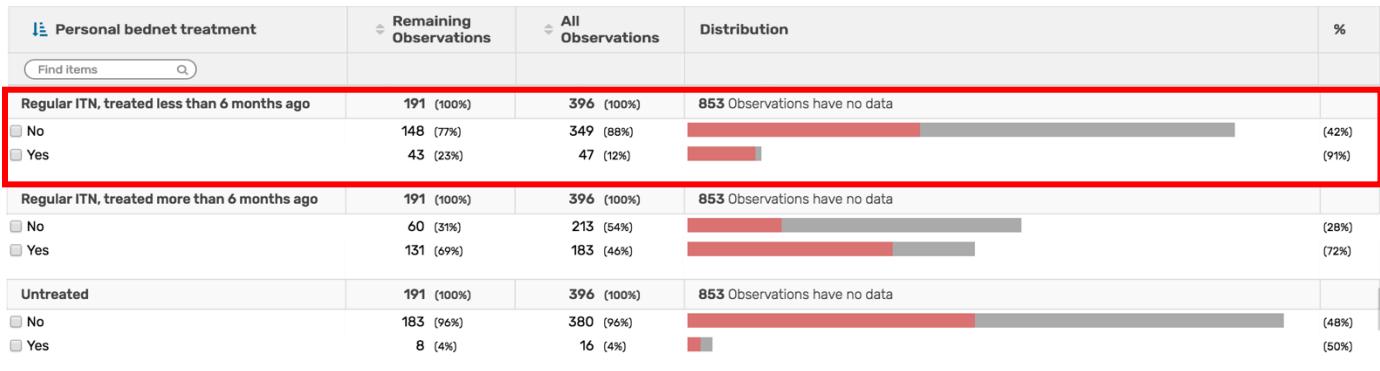


<sup>4</sup> 191 participants are selected. The numbers in the Search Wizard are always based on the type of search conducted. In this case, we did a participant search, so it always shows participant numbers. The poverty line table provides numbers based on the number of household observations. Since multiple participants can belong to one household, you can select 56 household observations where the household falls below the poverty line, which corresponds to 191 participants who belong to those households.

<sup>5</sup> Personal bednet treatment under the Observations step. This variable was collected only at the first visit and records which types of bednets each participant uses.

<sup>6</sup> Select “All” for the top parameter, then select no for “Untreated” and no for “Regular ITN, treated more than 6 months ago”.

9. Note that the columns “Remaining observations” and “Observations” show different numbers. “Observations” indicates the total data, while “Remaining observations” indicates the filtered data, or participants from households below the poverty line. In the Distribution graphic, grey indicates the data that does not meet the filter criteria, while pink reflects the filtered data. The percentage on the far right tells you what percent of participants meet your filter criteria, or in this case, what percent of participants came from households below the poverty line. *What percent of participants who reported using a regular ITN, treated less than 6 months ago came from households below the poverty line?*<sup>7</sup>



Using the Shiny Apps to visualize data – Is a participant below the poverty line more likely to use a net?

In the last section, we noticed that 91% of participants who used regular ITN that were treated less than 6 months ago were also below the poverty line. On the same page we could also see that 72% of participants who used regular ITN treated more than 6 months ago were below the poverty line, and that 50% of participants who used untreated nets were below the poverty line. We hypothesize that participants below the poverty line are more likely to use a net than participants above the poverty line. We can test this hypothesis using one of the shiny apps, but first we need to reconfigure our filters.

- To remove the filters that you have applied, click on the green filter icon at the top of the Search Wizard. This will activate a pop-up that lists out all of the active filters you have. Now click the ‘x’ to remove the filter.

Select a Set of Participants (India ICEMR)

Observations

Active Filters

Households

Poverty line

<sup>7</sup> 91%. To find the answer, look at the row where data for “Regular ITN, treated less than 6 months” is “Yes”. We can see that there are 43 Remaining observations (below poverty line) out of 47 total observations. If we look all the way to the right, we can see the percent written for us so we don’t have to calculate it.

- Now select participants who use an untreated net or regular ITN. Make sure to set the top parameter to “any”:

**Personal bednet treatment**

This variable was only collected at the first visit.

Find Observations with **any** of the options selected below.

- The blue button at the end of the Search Wizard should now say “View 246 Participants”. Click on that button to go to the results page.



- To access the Shiny Apps, click on “Analyze Results”. We want to make a contingency table to look at the association between use of any bednet and poverty, so click on “Contingency Tables”

The screenshot shows the "246 Participants" results page. At the top, there are buttons for "Revise", "Combine with another search", "Save", and "Share". Below this is a "Participant Results" section with a "First 1 2 3 4 5 Next Last" navigation bar and an "Advanced Paging" button. A red box highlights the "Analyze Results" button. To the right, a modal window titled "246 Participants" is open. It has buttons for "Revise", "Combine with another search", "Save", and "Share". Inside the modal, there are three sections: "Distributions" (histograms), "Contingency Tables" (a 2x2 grid of tables), and "Data Summaries" (bar charts). Red arrows point from the "Analyze Results" button on the main page to the "Contingency Tables" section in the modal, and from the "Contingency Tables" section in the modal to the "Contingency Tables" section in the modal.

- When you open a Shiny App, it will have parameters you can edit. By default, the Shiny Apps work on all the data, not just the data you selected in the Search Wizard. However, you can choose to use your filtered data as a variable which compares the data you selected (participants who report using a bednet) to the rest of the data (participants who do not report

using a bednet). Which variable will you enter as your independent/exposure variable? Which variable will you enter as your dependent/outcome variable?<sup>8</sup>

- Set up your parameters as shown below. Click on the arrow to open the drop-down menu. You can either click through all of the variables by clicking on the arrows next to the labels or type the name of the variable in the box at the top.

**Independent/Exposure**

Poverty line

are / is

Below

**Dependent/Outcome**

Selected Participants

are / is

Selected

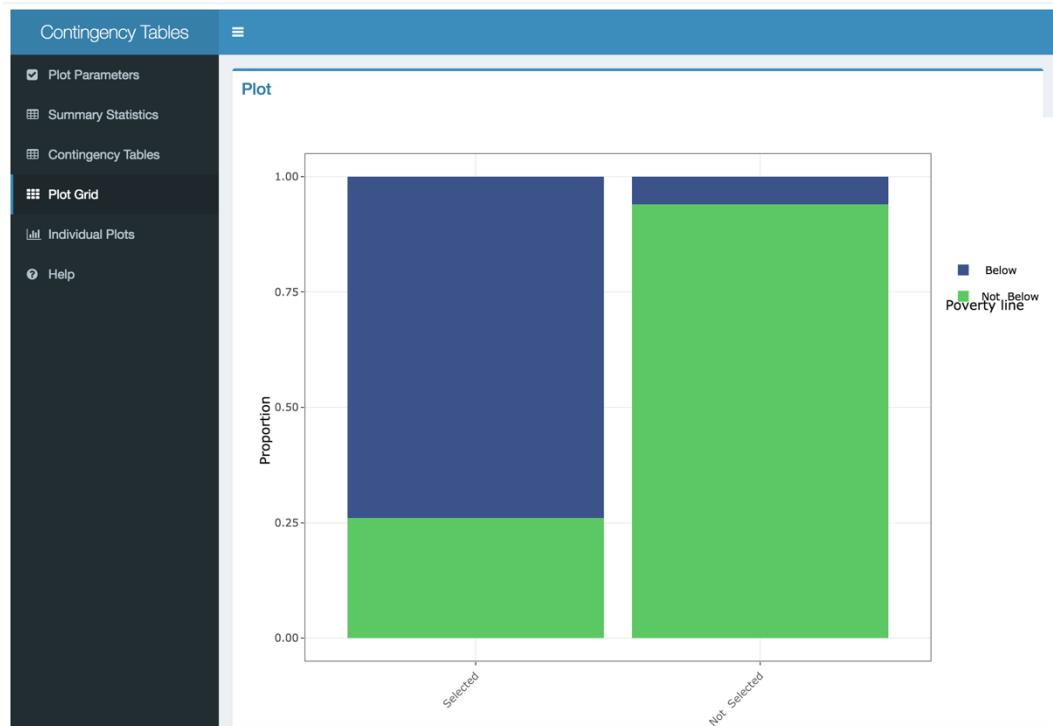
**Stratify Plot (1)**

None

**Stratify Plot (2)**

None

- Explore the different menu options on the left-hand side. Look at the Plot Grid, then at the Summary Statistics. What is the p-value for this analysis? What factors might you want to control for?<sup>9</sup>



<sup>8</sup> For this example, let's use "Poverty line" as the exposure variable and "Selected participants" as the outcome variable. Since we've filtered on participants who use bednets, we are really setting the outcome variable to bednet users.

<sup>9</sup>  $p < 0.0001$ . One obvious potential confounder is geographic region.

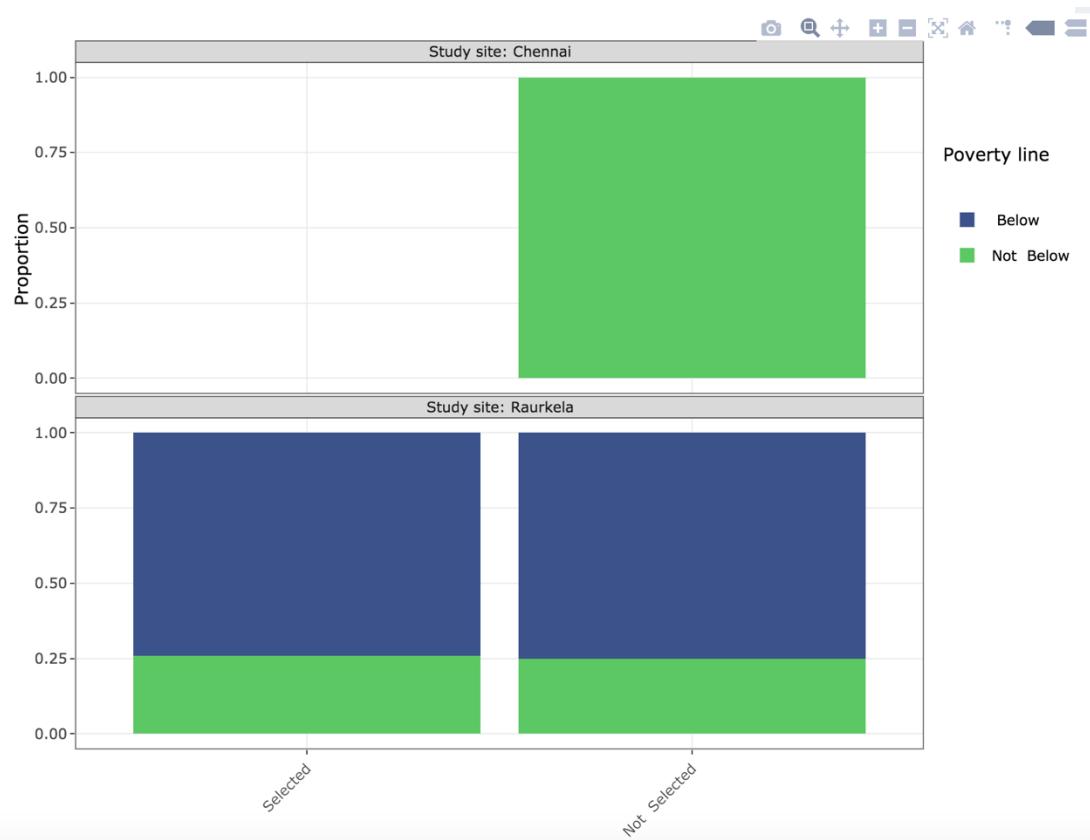
8. The cohort study was conducted in Rourkela and Chennai. A relationship between poverty and geographic region or bednet usage and geographic region might therefore confound results, so let's stratify our results based on study site. Click back to "Plot parameters" and update "Stratify Plot (1)".
9. Now look at the Plot grid again. We can see that the association we saw previously was completely dependent on geographic region. What other potential confounders do you want to explore?

**Stratify Plot (1)**

All possible

strata for

Study 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.

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](mailto:help@clinepidb.org).