



Exercise: My User Studies (part 2)

Exploring your uploaded dataset



In this exercise, we will explore the data we uploaded in Part 1 using features and tools on the ClinEpiDB platform.

1. Find the **Malaria toy data** dataset you uploaded in the **Studies** menu on the home page under **User studies**

2. Let's explore the data in the **Browse and Subset** tab

a. What percent of people had a positive blood smear?

b. How many people had a fever $\geq 38^\circ\text{C}$? _____

c. Of those with fever $\geq 38^\circ\text{C}$, what percent were loggers?

d. Does this look different from the percent of loggers in the total study population? _____

3. We hypothesize that sex impacts the risk of a positive blood smear and also impacts the risk of a positive flu test. Remove the temperature filter and head to the **Visualize** tab to test the hypotheses.

a. Which visualization tools will let you plot sex against blood smear or flu test?

b. Make a mosaic plot (RxC table) of sex and blood smear.

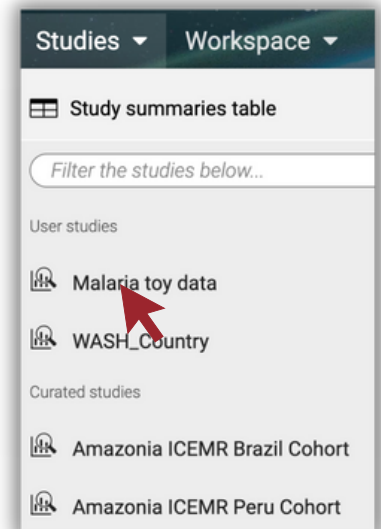
i. Why do the values for sex appear as 0 and 1? _____

ii. Does it look like there's a difference in blood smear positivity based on sex?

Navigate to the **Statistics** tab, what's the P value? _____

c. Copy the plot and update it to look at sex and flu

i. Roughly what proportion of people did not have a flu test done? _____



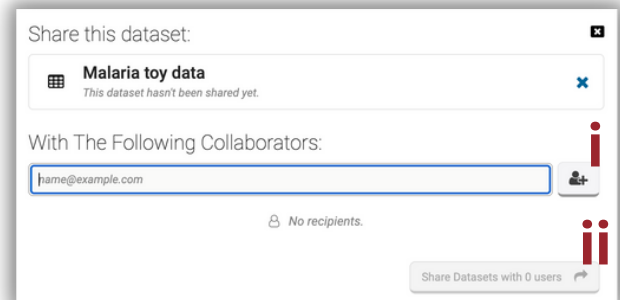
ii. Note that in this mock study, people with a positive blood smear for malaria were not tested for flu. Update the filters to select only participants who had a flu test performed (Indeterminate, Negative, Positive). What proportion of tests were positive for sex 0? _____ For sex 1? _____

iii. Are these significantly different? _____

4. Share the dataset with a colleague

- From the **Workspace** menu, click **My User Studies**
- On the **All** tab, select the checkbox next to the dataset you want to share
- Click the **Share Datasets** button
- Add the email address for the ClinEpiDB account you want to share with. Try with your neighbor or our workshop instructors (*dhelb@sas.upenn.edu*, *nkittur@uga.edu*, *droos@sas.upenn.edu*, *s.kelly@imperial.ac.uk*, *sheenass@sas.upenn.edu*, *weilus@sas.upenn.edu*)

- Add an email address and click the person icon to the right. Continue to add email addresses, as desired
- Click the **Share Datasets with X users** button



- See who you have shared a dataset with (**Shared With**), and who has shared a dataset with you (**Owner**), on the **All** tab
- Delete a dataset (for you and anyone you shared it with) at any time by selecting the checkbox next to the study, then clicking the **Remove** button

<div> All New upload Recent uploads Help </div> <div> <input type="text" value="Search Datasets"/> Showing 2 of 3 data sets <input checked="" type="checkbox"/> Only show data sets related to ClinEpiDB 69.92 K (0%) of 10.00 G used 1 row selected. Clear selection. <div> Share Datasets Remove </div> </div>											
<input type="checkbox"/>	Name / ID	Summary	Type	VEuPathDB Websites	Status	Owner	Shared With	Created	File Count	Size	Quota Usage
<input checked="" type="checkbox"/>	Malaria toy data (4067712)	Toy dataset of malaria and flu in 5-40 year olds	ISA Simple (0.0)	ClinEpiDB	✓	Me	Danica Helb, David S Roos, weilu song, Sarah Kelly, Nupur Kittur	2 hours ago	1	64.46 K	0%
<input type="checkbox"/>	WASH_Country (4063779)	WASH data in most populous countries	ISA Simple (0.0)	ClinEpiDB	✓	Nupur Kittur		1 month ago	1	1.31 K	

Turn to the next page for answers to this exercise!



2. Browse and subset

- a. What percent of people had a positive blood smear? **29%**
- b. How many people had a fever $\geq 38^\circ\text{C}$? **629** *Hint: subset the data from 38 - 40.4 and look at the count in the dataset diagram*
- c. Of those with fever $\geq 38^\circ\text{C}$, what percent were loggers? **10%**
- d. Does this look different from the percent of loggers in the total study population? **Not really, loggers make up 9% of the total study population and 10% of the study population with fever.**

3. Visualize

- a. Which visualization tools will let you plot sex against blood smear or flu test? **Bar plot (using overlays or facets), Mosaic plot (RxC table)**
- b. Make a mosaic plot (RxC table) of sex and blood smear.
 - i. Why do the values for sex appear as 0 and 1? **The data we uploaded had values of 0 and 1 for sex**
 - ii. Does it look like there's a difference in blood smear positivity based on sex? Navigate to the Statistics tab, what's the P value? **No major difference based on sex. P value is 0.2 using a Chi-squared test.**
- c. Copy the plot and update it to look at sex and flu
 - i. Roughly what proportion of people did not have a flu test done? **~28%**
 - ii. Note that in this toy study, people with a positive blood smear for malaria were not tested for flu. Update the filters to select only participants who had a flu test performed (Indeterminate, Negative, Positive). What proportion of tests were positive for sex 0? **11.2%** For sex 1? **14.1%**
 - iii. Are these significantly different? **No, p value of 0.5**