. CAUSAL INFERENCE
WINTER SEMESTER, 2023-24

As a researcher, you are interested in whether using mosquito nets decreases an individual's risk of contracting dengue. Suppose a dataset has been collected data from 1,752 households in a district (name not disclosed) from India. The dataset has variables related to environmental factors, individual health, and household characteristics. Additionally, for this district, there is a special government programme that provides free mosquito nets to households that meet specific requirements: to qualify for the program, there must be more than 4 members of the household, and the household's monthly income must be lower than Rs. 7000 a month. Households are not automatically enrolled in the programme, and thereby many chose not to use it. The data is not experimental—researchers have no control over who uses mosquito nets, and individual households make their own choices over whether to apply for free nets or buy their own nets, as well as whether they use the nets if they have them.

Here is the list and description of the variables in the dataset.

- Dengue risk (dengue_risk): The likelihood that someone in the household will be infected with dengue. Measured on a scale of 0–100, with higher values indicating higher risk.
- Mosquito net (net and net_num): A binary variable indicating if the household used mosquito nets.
- Mosquito net number (net_num): count of nets used in the house
- Income (income): The household's monthly income.
  Note: To get the right figure, you need to multiply with 10. Then you have income in Indian Rupees.
- Eligible for program (eligible): A binary variable indicating if the household is eligible for the free net program.
  Note: After making necessary adjustments for the income column (as mentioned in the last note), please verify the if the eligibility column is correct or not. If not, then make necessary adjustments following the programme criteria.
- Nighttime temperatures (temperature): The average temperature at night, in Celsius.
- Health (health): Self-reported healthiness in the household. Measured on a scale of 0–100, with higher values indicating better health.
- Number in household (household): Number of people living in the household.

- Insecticide resistance (resistance): Some strains of mosquitoes are more resistant to insecticide and thus pose a higher risk of infecting people with dengue. This is measured on a scale of 0–100, with higher values indicating higher resistance.

As stated earlier, the research question is whether using mosquito nets decreases an individual's risk of contracting dengue. So, your outcome variable is the risk of Dengue (dengue_risk). Your treatment variable is use of mosquito net; alternatively the number of mosquito net used.

1. How will you deal with your research question analytically to come up with a causal estimate of the treatment effectiveness?
    a. Does it matter if your treatment variable is binary (use of mosquito net: yes or no) or threatment is a count (like number of mosquito used in the household?
2. Make your assumptions clear how you will deal with those two scenarios above.
3. To progress with your analysis, use the following steps and give justification for each –
    a. Build a naive (and maybe incorrect correlation-is-not-causation) model to test the baseline relationship between net-use and degue-risk. How is net-use (alternatively, number of nets used) associated with dengue-risk? Why is this estimate not causal?
    (Hint 1: the presence of government is an extra source of problem)
    b. It is useful to draw a DAG. (think carefully before drawing it)
    c. Using the DAG, determine the minimum sufficient adjustment set. What nodes need to be adjusted for to ensure the pathway between net-use and risk of dengue is identified?
    d. Use **multiple regression** to close the back-doors. Include the variables from your adjustment set as explanatory variables in a regression model.
    e. Use **matching** to close the backdoors.
    (Hint 2: Use all matching techniques you know. There will be differences in outcome for each procedure. So justify your choice steps, for example: use of propensity score, inverse probability weighting or any other method.
    f. Post matching, use the variables from your adjustment set to generate propensity scores for assignment to treatment, then create inverse probability weights and use those weights in a regression model.)
    g. How do these adjusted effects compare to the naive model? How confident would you be claiming that these are causal effects? Why? What could you do to improve your causal identification?
4. Are you conviced that with whatever you have just done is resulting in a causal effect? Give reasons.
5. BONUS Q: Now suppose, the question is changed to evaluation of the mosquito net programme by government. Is it a completely different question? How will you go about it?
    (Hint: You give your reasons, or modify the last strategy you used. Want to know your thought process here.)

Instructions:

1. You may use STATA or R for this analysis, as per your convenience
2. You may use .rmd (R-Markdown, knitted to pdf) or Word file to submit your answer. (ONE FILE ONLY)
3. The above document must contain both your codes, the results (tables, graphs etc.) and your interpretation of results. The interpretation part is most important for evaluating your paper.
4. ***Do not copy the codes, tables or interpretations from others. It's easy to find plagiarism, please refrain from doing that.***
5. Google-form link will be posted in Piazza in due course. ***Link will have strict expiration schedule, beyond which no submissions will be possible. In case of no/delayed submission no marks will be given.***

All the best.