Problem 2: Evaluating the Impact of Bed Net Distribution on Child Health

You have been commissioned by a national Ministry of Health to evaluate the impact of a large-scale anti-malaria campaign implemented in the country between 2004 and 2007. After some long hours of searching for adequate data, you finally manage to put together a data set that contains information on child fever (a common symptom of malaria) before and after the anti-malaria campaign. The data are from two nationally representative Demographic and Health Surveys: a first round in 2001, and a second round in 2007, and contain an indicator for whether a child had fever in the two weeks preceding the survey. You also obtain from the Ministry of Health information on which districts received the anti-malaria intervention.

Download the data set called **PS3_DID.dta** from the course website and take a careful look at the data. Rename variables as appropriate. Familiarize yourself with the data set by tabulating all variables of interest by year.

a. Generate indicator variables for "poor water source" and "no toilet," using your best judgment to recode the categories (there is no right way of doing this, just use common sense).
b. Generate a table that shows fever prevalence by survey year – describe the changes in fever probabilities. How would you describe the overall trend (differences)?
c. Estimate the following pooled cross-section model:

$$\text{fever}_{ijt} = \alpha + \beta \cdot \text{treat\_dist}_j + \varepsilon_{ijt}$$

for child i in district j and year t. The variable treat_dist_j is an indicator for whether district j received the bednets intervention (= 1) or not (= 0). Make sure to cluster your standard errors at the district level. How would you interpret the coefficient on treat_dist_j?

d. Generate a $post_t$ variable, which equals 1 if the survey year is 2007 and 0 otherwise, and include it in the regression as follows:

$$\text{fever}_{ijt} = \alpha + \beta \cdot \text{treat\_dist}_j + \gamma \cdot post_t + \varepsilon_{ijt}$$

Interpret the coefficient on $post_t$. How does the coefficient on treat_dist_j change? Is this what you expected? Now, how do you interpret the coefficient on "treat_dist_j? Is this a difference-in-differences (DID) estimate?

e. Now estimate the following model:

$$\text{fever}_{ijt} = \alpha + \beta \cdot \text{treat\_dist}_j + \gamma \cdot post_t + \delta \cdot \left(post_t \times \text{treat\_dist}_j\right) + \varepsilon_{ijt}$$

where $post_t \times \text{treat\_dist}_j$ is the interaction term between the post and the treatment district indicators. Interpret each of the four estimated coefficients: α, β, γ and δ. Is this now the correct DID regression? Which coefficient is the DID coefficient (estimated treatment effect)?

f. Now re-estimate the model from part e), but this time include controls for bad water source, poor sanitation, child age (each age group as separate variable) and other household characteristics. Discuss the point estimates you get on these additional covariates. Does the DID coefficient change? Why might you expect it to change? Under what conditions would you not expect it to change?
g. After presenting your results, a local NGO informs you that they had run their own anti-malaria campaign in districts NOT targeted by the national program. How would that affect the estimates you got?
h. Malaria is transmitted by mosquitoes, which breed in standing water. Suppose there was a nationwide drought in 2006/2007, and the mosquito population was unable to replicate at the usual rate. How might this affect the interpretation of your results?