Vellore Model

**Model 1:**

For every child in MAL-ED study, build a model.

Outcome:

Y1 = Illness Rate (times/year)

Y2/Y3 = Infection Rate (times/total study test) by bacteria and virus

Covariates:

X1 = The closest bathing water E. coli concentration

X2 = The closest drain water E. coli concentration

X3 = The closest child hand rinse E. coli concentration

X4 = The closest piped water E. coli concentration

X5 = The closest particulate E. coli concentration

X6 = The closest produce E. coli concentration

X7 = The closest pubic latrine swab E. coli concentration

X8 = The closest toy feeding spoon rinse E. coli concentration

X9 = drain length within 1 km

X10 = road length within 1 km

X11 = the distance from the closest open defecation site

X12 = number of children within 1 km

X13 = distance to the closest drain

X14 = distance to the closest road

May consider the effective range: if the sample is too far away, use missing value or average value of concentration of that sample type.

Logistic regression

Model 2:

For every study household in SaniPath study, build a model.

Outcome:

Y1 = Average Illness Rate (times/year) within 1 km

Y2/Y3 = Average Infection Rate (times/total study test) by bacteria and virus within 1 km

Covariates:

X1 = The average bathing water E. coli concentration within 1 km

X2 = The average drain water E. coli concentration within 1 km

X3 = The average child hand rinse E. coli concentration within 1 km

X4 = The average piped water E. coli concentration within 1 km

X5 = The average particulate E. coli concentration within 1 km

X6 = The average produce E. coli concentration within 1 km

X7 = The average swab E. coli concentration within 1 km

X8 = The average toy feeding spoon rinse E. coli concentration within 1 km

X9 = drain length within 1 km

X10 = road length within 1 km

X11 = the distance from the closest open defecation site

X12 = number of children within 1 km

X13 = distance to the closest drain

X14 = distance to the closest road