# Simulation Studies for Dengue Transmission and Within-Host Immune Responses

Alexander Kirpich[1], Yang Yang[1], Ira Longini[1], and M. Elizabeth Halloran[2]

[1]Department of Biostatistics, University of Florida
[2]Fred Hutchinson Cancer Research Center

**Abstract**

Dengue is a neglected tropical infectious disease transmitted between mosquitoes and human beings. There are four known serotypes cocirculating with similar clinical symptoms. Infection with one serotype provides permanent immunity against it. Secondary or tertiary infections lead to more severe diseases than primary ones. Serotype-specific laboratory testing was not available until recently. Tests were only performed on infected people with symptoms, leaving asymptomatic infections undetected. We propose to simulate dengue epidemics together with immune responses. In this simulation study infection outcomes are observable only during a hypothetical study period and only on those randomly selected symptomatic infections. A cohort study data conducted from 2004 to 2010 in Nicaragua is a motivating data set. We investigate quality of the inference based on the amount of known full infection histories in the sample.

**Key Words:** Dengue; Serotype-specific; Infection history; Incomplete history; Simulation

## 1. Intorduction

Dengue is an infectious disease transmitted between mosquitoes and humans. Currently four serotypes cocirculate with the dominant serotype changing year to year. Secondary dengue infections often lead to more severe disease than primary infections. Unfortunately pathogenicity and immunology of dengue viruses are complex and not clearly understood and available data are often limited.

Scientific goals of this work are:

- To explore how previous infections affect chance of future infections with other serotypes and risk of disease progression.
- To evaluate quality of inference for the case when the population entering the study is assumed to be naive.
- To evaluate the quality of inference for the cases when complete serotype infection history is unknown.

The main challenges for Nicaragua are:

- Lack of reliable diagnostic testing of most cases in collected data, e.g., Enzyme-linked immunosorbent assay (ELISA) indicates only infection, but not serotype.
- Population in the study was not naive at baseline.

## 2. Methods

Likelihood approach was utilized in the analysis. Let $p_{0tv}$ be the baseline probabilities of infection for year $t$, and serotype $v$. We assume

$$\text{logit}(p_{itv}) = \text{logit}(p_{0tv}) + \boldsymbol{X}_i^\tau \boldsymbol{\beta}_{inf}$$

and $p_{itv}$ are the effective probabilities for individual $i$, where covariate vector $\boldsymbol{X}_i$ has exposure history and acquired immunity.

**Likelihood for infection outcomes** is assumed to have the form

$$P[\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{p}_0, \boldsymbol{\beta}] =$$

$$\prod_{t=1}^{T}(q_{it})^{1-Y_{it*}} \times \left\{ (1-q_{it}) \prod_{v=1}^{V} \left[ \frac{\log(1-p_{itv})}{\sum_{u=1}^{V}\log(1-p_{itu})} \right]^{Y_{itv}} \right\}^{Y_{it*}}$$

Where

- $\boldsymbol{Y}_i = \{Y_{itv} : \ t = 1, \ldots, T, \ v = 1, \ldots, 4\}$.
- $q_{it} = \prod_{v=1}^{V}(1-p_{itv})$.
- $Y_{it*} = \max\{Y_{itv} : v = 1, \ldots, 4\}$ is an indicator of infection with any serotype.

**Likelihood for pathogenicity (disease outcome.)** Let $\phi_{0v}$ be the baseline probability of disease given infection with serotype $v$. The effective probability of disease in infection year $\tilde{t}_{ik}$ is given by

$$\text{logit}(\phi_{ik}) = \sum_{v=1}^{4} \mathbf{1}_{\{V_{ik}=v\}} \Big[ \text{logit}(\phi_{0v}) + \boldsymbol{X}_i(\tilde{t}_{ik})^\tau \boldsymbol{\beta}_{pat} \Big].$$

The probability of disease outcomes is given by

$$\mathbb{P}\big[S_i(t)|\boldsymbol{X}_i^S(t), \boldsymbol{\phi}, \boldsymbol{\beta}_{pat}\big] =$$

$$\begin{cases} \phi_{ik}^{S_i(t)}(1-\phi_{ik})^{1-S_i(t)}, & t = \tilde{t}_{ik}, \ k = 1,2,3,4, \\ 1 - S_i(t), & \text{otherwise.} \end{cases}$$

**Likelihood for immune responses.** Immune responses were generated from the normal distribution upon infection, with mean and variance depending on exposure history. After that, the responses follow a mean trajectory of a general logistic function, with an $AR(1)$ error term. For the parameters to be identifiable, one may have to assume all serotypes share part of the parameters.

## 3. Simulation settings

Epidemics of four dengue serotypes are simulated in a population of 1000 individuals over 11 years. Complete 11 year case with full infection histories and years 5-11 with full infection histories are considered as the study periods. Also, for years 5-11 in addition to complete infection histories, incomplete histories are considered. Samples are censored with only $25\%$, $50\%$ and $75\%$ of individuals having full serotype histories and the remaining ones having only the indication of infection with unknown serotype ($Y_{it*}$). We assume all individuals are naive to dengue before the study period.

We include five covariates that affect the infection probabilities in the model:

- one year after primary infection;
- $\geq 2$ years after primary infection;
- one year after a non primary infection;
- $\geq 2$ years after a non primary infection;
- Time between two most recent infections.

For now, immunology and pathogenicity components are not included in the simulation. For the simulations **two different scenarios** of baseline infection probabilities are considered:

- $p_{0tv}$'s are serotype-specific but time-independent.
- $p_{0tv}$'s are serotype-specific and time-dependent.

For incomplete histories time dependence was not considered because of lack of reliability of the estimates for incomplete histories.

## 4. Current results

For the complete infection histories considering years 5-11 as the realistic approximation to Nicaragua data and ignoring all data before the baseline moderate underestimation was observed. Parameter values and estimates are summarized below.

| **Names** | $p_{0t1}$ | $p_{0t2}$ | $p_{0t3}$ | $p_{0t4}$ |
|---|---|---|---|---|
| Truth | 0.020 | 0.050 | 0.040 | 0.100 |
| Mean | 0.017 | 0.039 | 0.030 | 0.060 |
| St. Dev | 0.001 | 0.003 | 0.003 | 0.003 |

| **Names** | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ |
|---|---|---|---|---|---|
| Truth | $-1.00$ | $-0.50$ | $-1.50$ | $-0.75$ | $-0.75$ |
| Mean | $-1.12$ | $-0.58$ | $-2.12$ | $-2.49$ | $-0.06$ |
| St. Dev | 0.227 | 0.172 | 1.921 | 3.669 | 0.498 |

Estimates are still feasible for small probabilities and not that feasible for large probabilities because of the acquired immunity. Estimates for $\beta_3$ and $\beta_4$ (secondary infection) are unreliable.

As a reference, unrealistic complete history analysis for 11 years is summarized below

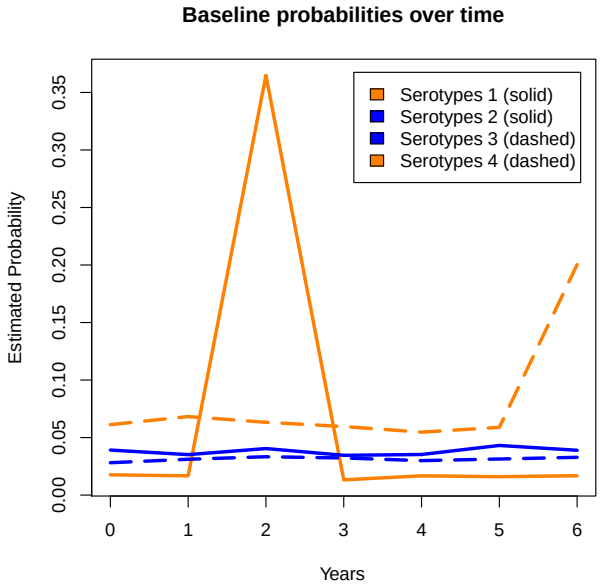
| **Names** | $p_{0t1}$ | $p_{0t2}$ | $p_{0t3}$ | $p_{0t4}$ |
|---|---|---|---|---|
| Truth | 0.020 | 0.050 | 0.040 | 0.100 |
| Mean | 0.020 | 0.051 | 0.040 | 0.100 |
| St. Dev. | 0.002 | 0.002 | 0.003 | 0.005 |

| **Names** | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ |
|---|---|---|---|---|---|
| Truth | $-1.00$ | $-0.50$ | $-1.50$ | $-0.75$ | $-0.75$ |
| Mean | $-1.06$ | $-0.49$ | $-1.53$ | $-0.66$ | $-0.27$ |
| St. Dev. | 0.115 | 0.146 | 0.300 | 0.329 | 0.190 |

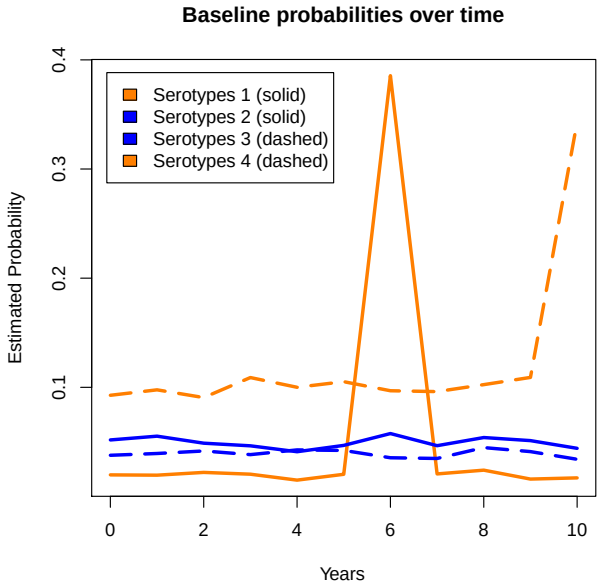
For the second scheme, probabilities are serotype and time dependent and "peaks" in the epidemic are added by setting $p_{021} = 0.4$ and $p_{064} = 0.35$. Other parameter values were kept the same as before.

Considering years 5-11 as the realistic approximation to the Nicaragua data and ignoring all data before the baseline, moderate underestimation was still observed. "Peak" probability values $p_{021}$ and $p_{064}$ are underestimated, especially $p_{064}$.

Parameter estimates for serotypes over time are summarized in the graph below.

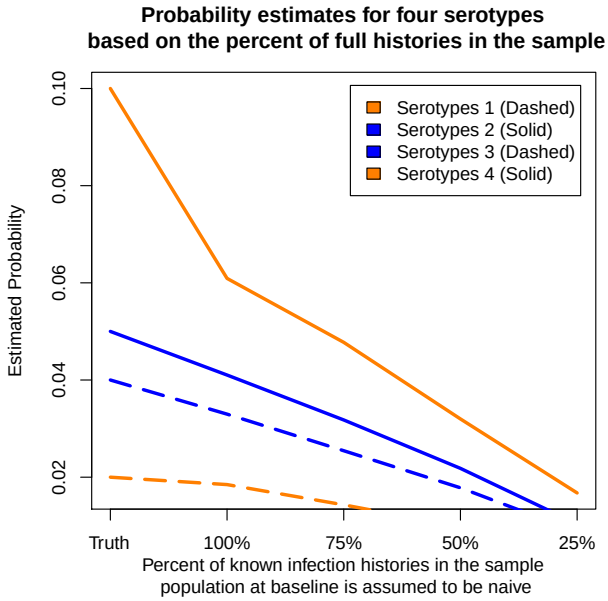



As a reference, unrealistic complete case analysis for 11 years is summarized on the graph below with the same peaks $p_{021} = 0.4$ and $p_{064} = 0.35$. Probability estimates for $p_{021}$ and $p_{064}$ now look reasonable.




For incomplete serotype history the values of the estimates are severely biased. This happens because we fail to account for non-naive baseline infection histories of the individuals who are entering the study. For high infection probabilities among serotypes this

causes severe bias as well as a loss of a large amount of infection history. We can see that even when 75% of individuals have complete serotype histories. Results for the first scenario and incomplete serotype histories are summarized in the graph below.




## 5. Ongoing Research

This is an ongoing project. The following are a few extensions and enhancements of the existing model and analyses to be implemented:

- The inference will include pathogenicity and immunology components.
- The analysis will include likelihood inference for incomplete data **accounting for non-naive baseline** with known infection status but missing serotype. The goal is to find the minimum proportion of complete cases required for meaningful inference.
- An alternative for comparison will be a Bayesian MCMC approach.
- Use simulations to evaluate potential impact of a current vaccine candidate not effective against dengue serotype 2 in a previous study.

Project funding: Dengue Vaccine Initiative; NIGMS/NIH U01-GM070749, NIAID/NIH R37-AI032042.

## REFERENCES

Yang, Y, Halloran, ME, Daniels M and Longini, IM. *Modeling Competing Infectious Pathogens from a Bayesian Perspective: Application to Influenza Studies with Incomplete Laboratory Results.* Journal of the American Statistical Association. 2010; 105:1310-1322.

Kyle JL, Harris E. *Global spread and persistence of dengue.* Annu Rev Microbiol. 2008;62:71-92. doi: 10.1146/annurev.micro.62.081307.163005