# Supporting Information to Estimating the Initial Susceptible Fraction in Dengue Dynamics"

Flavio Codeço Coelho

&

Luiz Max de Carvalho

## A remark on prior distributions and tail behavior of the distribution of $R_t$

There are a number of approaches to deriving the distribution of $R_t$ . Alternatively to the approach described in the main text (Ederer and Mantel, 1974), one could use the conditional distribution of $R_t$ on $Y_{t+1}$ and $Y_t$ as defined in equation A7 of Nishiura et al. (2010):

$$f_R(R_t) = (Y_t R_t)^{Y_{t+1}} e^{-Y_t R_t} \tag{1}$$

Noticing the kernel of (1) is that of a gamma distribution with $a_2 = Y_{t+1}+1$ and $b_2 = Y_t$, we obtain a proper density from which to construct $c_\alpha(R_t)$, simply by computing the appropriate quantiles of said distribution. This density is

$$f_N(R_t|a_2, b_2) = \frac{b_2^{a_2}}{\Gamma(a_2)} R_t^{a_2-1} e^{-b_2 R_t} \tag{2}$$

In order to decide which approach to take, it may be of use analyzing the variance and tail behavior of the derived distributions for $R_t$ . Consider the case of using a flat $Uniform(0, 1)$ prior for $\theta_t$. With $a_0 = b_0 = 1$, $a_1 = a_2$ and $b_1 = b_2 + 1$. The beta prime (inverse beta distribution) will have heavier tails compared to the conditional distribution proposed by Nishiura et al. (2010), thus providing more conservative confidence/credibility intervals. To see that one needs simply take the ratio of the Beta prime and Gamma (unnormalized) densities and evaluate the limit as $R_t$ goes to infinity:

$$\lim_{R_t\to\infty} \frac{f_P(R_t|a_1, b_1)}{f_N(R_t|a_2, b_2)} = \lim_{R_t\to\infty} \frac{e^{Y_t R_t}}{(1 + R_t)^{Y_t+Y_{t+1}+2}} = \infty \tag{3}$$

As a side note, the Bayesian approach presented in this paper will give similar results to those of Wilson (1927) and Wilson (1927) for $Y_{t+1}$ and $Y_t >> 1$. Under the flat uniform prior for $\theta_t$, the Bayesian posterior credibility interval is nearly indistinguishable from the confidence interval proposed by Clopper and Pearson (1934) for $Y_{t+1}, Y_t > 20$. Note that the $Beta(1, 1)$ uniform prior for $\theta_t$ constitutes a poor prior choice mainly because the induced distribution for $R_t$ is only well-defined for $b_0 > 2$.

An advantage of the Bayesian approach is that one can devise prior distributions for $\theta_t$ taking advantage of the intuitive parameterization and flexibility of the beta family of distributions. Prior

elicitation can also be done for $R_t$ and the hyperparameters directly plugged into the prior for $\theta_t$. One can, for example, choose a priori mean and variance for $R_t$ and find $a_0$ and $b_0$ that satisfy those conditions. Let $m_0$ and $v_0$ be the prior expectation and variance for $R_t$. After some tedious algebra one finds

$$a_0 = \frac{m_0 v_0 + m_0^3 + m_0^2}{v_0} \tag{4}$$

$$b_0 = \frac{2v_0 + m_0^2 + m_0}{v_0} \tag{5}$$

If one wants only to specify $m_0$ and a coefficient of variation $c$ [1] for $R_t$ *a priori*, some less boring algebra gives:

$$a_0 = \frac{m_0^3 c^2 + m_0^3 + m_0^2}{m_0^2 c^2} \tag{6}$$

$$b_0 = \frac{2m_0^2 c^2 + m^2 + m}{m_0^2 c^2} \tag{7}$$

This approach thus makes possible to incorporate epidemiological knowledge about disease biology (e.g. the magnitude of $R_0$) into the computation of $R_t$ . This may prove particularly important when disease counts are low and/or close to the detection threshold. We provide an R script to perform the above elicitation at https://github.com/fccoelho/paperLM1/blob/master/aux/elicit_Rt_prior.R.

[1] $c = \sqrt{v_0}/m_0$.

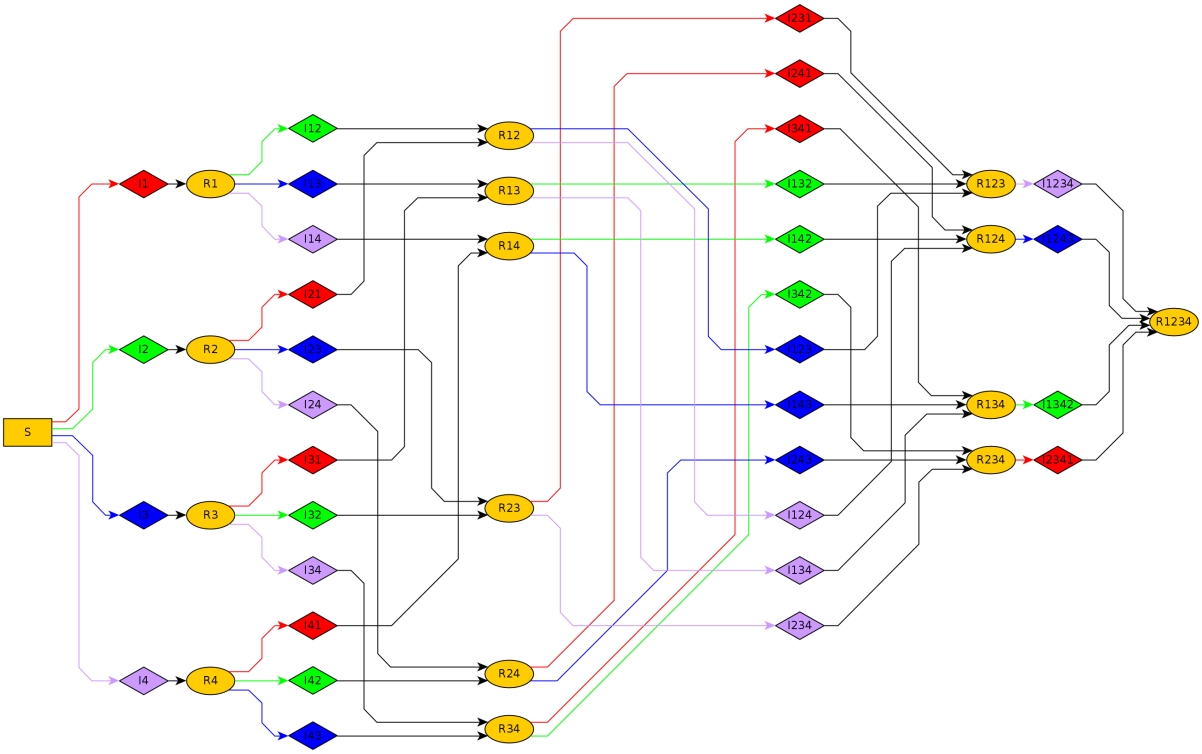


Figure 1: Block diagram detailing the stochastic model. Infected individuals with different Dengue viruses are represented by different colors. Infections are also represented by colored arrows matching the virus type.

# Multi-strain dynamics

In order to validate the estimates derived in the paper, we simulated multi-strain Dengue dynamics by means of a stochastic 4-serotype SIR model with cross-immunity as described below. The model permits up to 4 dengue infections with reduced susceptibility after the first Dengue episode due to cross immunity. Immunity to each serotype is considered complete and permanent. Figure xx depicts all possible states and state-transitions included in the model.

Let $S$ be individuals susceptible to all 4 types of dengue, $I_i$ infectious with Dengue type $i$ and $R_i$ individuals recovered from Dengue type $i$. Infectious individuals already on their secondary and later Dengue infections are represented by multiple indices. For example, $I_{[23]1}$ is an individual which has had Dengues type 2 and 3 in the past – and therefore is immune to them – and is currently transmitting Dengue 1. The index outside the bracket denotes current infection. Recovered individuals indices denote their immunity, so for instance $R_{123}$ is an individual which is immune to Dengue types 1, 2 and 3, but not to 4. Let $I_{*i} = \sum I_{[\ldots]i}$ with $[\ldots]$ representing exposure history of the infected individual which can vary from 0 to 3 in length. All individuals are born to the $S$ state and birth and death rates are equal.

The possible state-transitions and their propensities are listed in table 1.

The model is implemented as a continuous time Markov jump process. Let

$$\overrightarrow{X}(t) = [S(t), I_1(t), I_2(t), \ldots, R_{1234}(t)]$$

be the state of the system at the time $t$. The system is written as a forward Kolmogorov differential

Table 1: **State-transitions and propensities: $P(\Delta X(t)|X(t))$. The transitions are summarized below. Fully expanded, the system contemplates 64 possible state transitions, as can be verified in figure 1.**

| Transition | Propensity | State Change | Description |
|---|---|---|---|
| $S \to I_i$ | $\beta S I_{*i}$ | $\Delta S(t) = -1$, $\Delta I_i(t) = 1$ | Primary infection |
| $I_i \to R_i$ | $\sigma I_i$ | $\Delta I_i(t) = -1$, $\Delta R_i(t) = 1$ | Primary recovery |
| $R_i \to I_{[i]j}$ | $\beta\delta R_i I_{*j}$ | $\Delta R_i(t) = -1$, $\Delta I_{[i]j}(t) = 1$ | Secondary infection |
| $I_{[i]j} \to R_{ij}$ | $\sigma I_{[i]j}$ | $\Delta I_{[i]j}(t) = -1$, $\Delta R_{ij}(t) = 1$ | Secondary recovery |
| $R_{ij} \to I_{[ij]k}$ | $\beta\delta R_{ij} I_{*k}$ | $\Delta R_{ij}(t) = -1$, $\Delta I_{[ij]k}(t) = 1$ | Tertiary infection |
| $I_{[ij]k} \to R_{ijk}$ | $\sigma I_{[ij]k}$ | $\Delta I_{[ij]k}(t) = -1$, $\Delta R_{ijk}(t) = 1$ | Tertiary recovery |
| $R_{ijk} \to I_{[ijk]l}$ | $\beta\delta R_{ijk} I_{*l}$ | $\Delta R_{ijk}(t) = -1$, $\Delta I_{[ijk]l}(t) = 1$ | Quaternary infection |
| $I_{[ijk]l} \to R_{ijkl}$ | $\sigma I_{[ijk]l}$ | $\Delta I_{[ijk]l}(t) = -1$, $\Delta R_{ijkl}(t) = 1$ | Quaternary recovery |
| $\to S$ | $\mu N$ | $\Delta S = 1$ | Birth |
| $All \to$ | $\mu N$ | $\Delta All = -1$ | Death |

equation, which in matrix form looks like

$$\frac{dP_X(t)}{dt} = QP_X(t) \tag{8}$$

Where $P_X(t)$ is the matrix of transition probabilities (given in table 1) and Q is the generator matrix, whose non-zero values are also given in table 1 (in the state change column). The full formula and matrices are ommited due to their large sizes.

# References


Clopper, C. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, pages 404–413.

Ederer, F. and Mantel, N. (1974). Confidence limits on the ratio of two poisson variables. *American Journal of Epidemiology*, 100(3):165–167.

Nishiura, H., Chowell, G., Heesterbeek, H., and Wallinga, J. (2010). The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course. *J R Soc Interface*, 7(43):297–307.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212.