# NEW FRONTIERS IN APPLIED PROBABILITY

A Festschrift for SØREN ASMUSSEN
Edited by P. GLYNN, T. MIKOSCH and T. ROLSKI

Part 5. Stochastic growth and branching

## THE SIZE OF A MAJOR EPIDEMIC OF A VECTOR-BORNE DISEASE

DARYL J. DALEY, *The Australian National University and The University of Melbourne*

Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia. Email address: daryl.daley@anu.edu.au

RANDALL J. SWIFT, *California State Polytechnic University, Pomona*

Department of Mathematics, California State Polytechnic University, Pomona, CA 91768, USA. Email address: rjswift@csupomona.edu

APPLIED PROBABILITY TRUST
AUGUST 2011

# THE SIZE OF A MAJOR EPIDEMIC OF A VECTOR-BORNE DISEASE

By DARYL J. DALEY AND RANDALL J. SWIFT

## Abstract

Based on a simple model due to Dietz, it is shown that the size of a major epidemic of a vector-borne disease with basic reproduction ratio $R_0 > 1$ is dominated by the size of a standard SIR (susceptible–infected–removed) epidemic with direct host-to-host transmission of disease and the same $R_0$. Further bounds and numerical illustrations are provided, broadly spanning situations where the size of the epidemic is short of infecting almost all those susceptible. The total size is moderately sensitive to changes in the population parameters that contribute to $R_0$, so that the fluctuating behaviour in 'annual' epidemics is not surprising.

*Keywords:* Epidemic; dengue; basic reproduction ratio; epidemic total size; branching process; deterministic epidemic model; seasonal infection

2010 Mathematics Subject Classification: Primary 92D30
Secondary 62P10

## Preamble

My (DJD's) first contact with Søren Asmussen was indirect, through a note that I had written years earlier establishing the criticality condition for a simple two-sex branching process modelled in 1967 much as I heard it from Geoff Watterson while visiting Melbourne University. Søren, then a student in Göteborg, discussed the criticality condition for related models by a much simpler martingale argument, and some years later he came on an extended visit to Canberra. Now an epidemic model can be construed as a branching process that takes place *on* a pre-existing population, as distinct from a biological population that may grow. Models for epidemic processes are arguably better formulated within a population setting, where at the micro-level we describe what actions and interactions may occur as affecting individual members of the population. Analysis of the model usually describes aggregate behaviour over many individuals, and the stochastic element is largely confined to the model construction stage. The ensuing discussion is no exception.

## 1. Introduction

We have recently considered data on annual outbreaks of dengue fever (DF) in Singapore (see, e.g. Ooi *et al.* (2006) and, for a brief Australian perspective, my discussion (Daley (2010)) of Gani's (2010) Knibbs lecture). Worldwide, dengue affects between 50 and 100 million people a year. It passes unnoticed by the vast majority of people who are infected by it, and they are subsequently immune to further attacks by that one of the four dengue viruses that infected them, but its longer term effects can be more serious: subsequent attacks by any of the other three dengue viruses can lead to the more serious dengue haemorrhagic fever or dengue

shock syndrome. Gubler and Kuno (1997) gave a comprehensive survey of what was then known of DF. The disease is spread chiefly by the *Aedes aegypti* female mosquito; for the sake of simplicity, we regard the disease as being spread by a single virus. After ingesting blood from an infected human, the virus develops in its mosquito host and makes its way to the salivary tract whence it is subsequently injected into another human when the mosquito is successful in taking another blood-meal prior to laying eggs. A generation cycle, from ingestion from one infectious human until another human has become infectious via the mosquito biting them both, is from two to three weeks' duration.

A simple description of the population dynamics for DF, or indeed, for several similar diseases involving vector-borne transmission, is given in Dietz (1974). A careful mathematical analysis of the differential equations describing a related model is given in Esteva and Varga (1998). The purpose of this note is to describe how the total size of a major outbreak of the disease ('epidemic' in common parlance) can be determined for the model. What we show is that, for many such outbreaks, the size is little different from that of a related standard 'general epidemic model' in a closed homogeneously mixing population (which is often described as an SIR compartmental model, denoting the three categories of susceptible, infectious, and removed cases).

## 2. The model

Let $X_j(t)$ and $Y_j(t)$ denote the respective numbers of susceptibles and infectives at time $t$ in the populations of humans ($j = 1$) and female mosquitoes ($j = 2$). For our purposes, as in Dietz (1974) and Esteva and Vargas (1998), we consider infection as occurring from just the one virus. We regard the total size of the mosquito population as being capped, whether naturally so or by a control programme (e.g. via larval source reduction in Singapore); it is maintained by an influx of susceptible mosquitoes at constant rate $A$ and a *per capita* death rate $\mu_2$, whether susceptible or infectious. Then

$$\dot{X}_2 + \dot{Y}_2 = A - \mu_2(X_2 + Y_2);$$

consequently,

$$X_2(t) + Y_2(t) = \frac{A}{\mu_2} + \left(N_2(0) - \frac{A}{\mu_2}\right)e^{-\mu_2 t}$$

for some initial population size $N_2(0)$. For a steady-state total size, $\dot{X}_2 + \dot{Y}_2 = 0$ and $X_2 + Y_2 = A/\mu_2 = N_2(0)$.

Each adult female mosquito bites humans at a rate $\gamma$ say. If the human is DF infectious then transmission from human to mosquito may occur, with overall contact-and-transmission rate $\gamma_2$ say, while if the human is susceptible and the mosquito infectious, transmission from mosquito to human may occur at overall rate $\gamma_1$ say. Let $N$ denote the size of the human population where mosquitoes seek their blood meals, and assume that a mosquito successful in its quest finds a susceptible with probability $X_1/N$ and an infectious human with probability $Y_1/N$ (and some other, with probability $1 - (X_1 + Y_1)/N$, but these are of no concern). Thus, the total rates of infection transmission into the human and mosquito populations equal $\gamma_1(X_1/N)Y_2$ and $\gamma_2(Y_1/N)X_2$, respectively. Consequently, for the mosquitoes, we have

$$\dot{X}_2 = A - \gamma_2 \frac{Y_1}{N}X_2 - \mu_2 X_2, \qquad \dot{Y}_2 = \gamma_2 \frac{Y_1}{N}X_2 - \mu_2 Y_2. \qquad (2.1)$$

Recall that humans have lifetimes many times larger than a year, and that the epidemic that concerns us lasts maybe up to a dozen generations, still rather less than a year, so that we shall

regard the population as being of constant size $N$ for the year, i.e. we ignore human births and deaths. Infectious humans recover at a rate $\beta_1$ say, so that considering, in addition to $X_1$ and $Y_1$, the recovered (and thereafter immune) cases numbering $Z_1(t)$ say, we are led to the differential equations

$$\dot{X}_1 = -\gamma_1 \frac{X_1}{N} Y_2, \qquad \dot{Y}_1 = \gamma_1 \frac{X_1}{N} Y_2 - \beta_1 Y_1, \qquad \dot{Z}_1 = \beta_1 Y_1, \qquad (2.2)$$

so that $\dot{X}_1 + \dot{Y}_1 + \dot{Z}_1 = 0$ and, thus, $X_1 + Y_1 + Z_1 = $ constant over any time interval where (2.2) is a satisfactory total description of the evolution of the human population with mosquito contact. It is convenient to let $N_1$ denote this constant for the time interval, but note the distinction between $N_1$ here and $N$ earlier.

### 3. First analyses

Some information on the evolution of (2.1) and (2.2) follows by integration over the time interval $(0, t)$ say:

$$\log\left(\frac{X_1(0)}{X_1(t)}\right) = -\int_0^t \frac{1}{X_1(u)} \frac{dX_1(u)}{du} \, du = \frac{\gamma_1}{N} \int_0^t Y_2(u) \, du, \qquad (3.1a)$$

$$Z_1(t) - Z_1(0) = \beta_1 \int_0^t Y_1(u) \, du, \qquad (3.1b)$$

$$Y_2(t) - Y_2(0) = \frac{\gamma_2}{N} \int_0^t (N_2 - Y_2(u)) Y_1(u) \, du - \mu_2 \int_0^t Y_2(u) \, du$$
$$= \frac{\gamma_2 N_2}{N} \int_0^t Y_1(u) \, du - \mu_2 \int_0^t Y_2(u) \, du - \frac{\gamma_2}{N} \int_0^t Y_2(u) Y_1(u) \, du. \qquad (3.1c)$$

Substituting (3.1a) and (3.1b) into (3.1c) yields

$$Y_2(t) - Y_2(0) = \frac{\gamma_2 N_2}{\beta_1 N} [Z_1(t) - Z_1(0)] - \frac{\mu_2 N}{\gamma_1} \log\left(\frac{X_1(0)}{X_1(t)}\right) - \frac{\gamma_2}{N} \int_0^t Y_1(u) Y_2(u) \, du. \qquad (3.2)$$

In Sections 5 and 6 we give a more detailed analysis of the equations in (3.1); it suffices here to assert that when $Y_2(0) = Z_1(0) = 0$ and $Y_1(0) \ll X_1(0)$, the behaviour as $t \to \infty$ is such that $Y_i(t) \to 0$ for $i = 1, 2$ and the number of susceptibles surviving the outbreak, $X_1(\infty) \approx \theta X_1(0)$ say, which must equal $N_1 - Z_1(\infty)$, where $Z_1(\infty) \approx (1 - \theta) X_1(0)$, satisfies

$$-\log \theta = \frac{\gamma_1}{N} \int_0^\infty Y_2(u) \, du \quad \text{and} \quad Z_1(\infty) = \beta_1 \int_0^\infty Y_1(u) \, du.$$

Hence, from (3.2),

$$0 = \frac{\gamma_1 \gamma_2 N_2 X_1(0)}{\beta_1 \mu_2 N^2} (1 - \theta) + \log \theta - \frac{\gamma_1 \gamma_2}{\beta_1 \mu_2 N^2} \int_0^\infty Y_2(u) \beta_1 Y_1(u) \, du. \qquad (3.3)$$

Now express the two factors in the last integral as proportions $y_1(t) = Y_1(t)/X_1(0)$ and $y_2(t) = Y_2(t)/N_2$ of the initial numbers of susceptibles in their respective populations, since then $0 < y_j(t) < 1$ for all $t$ (we give more stringent bounds in Section 6). We also set

$$R_0 = \frac{\gamma_1 \gamma_2 X_1(0) N_2}{\beta_1 \mu_2 N^2}. \qquad (3.4)$$

The condition $R_0 > 1$ is necessary for a major epidemic, and we generally assume that it holds. Then (3.3) can be written as

$$1 - \theta + \frac{\log \theta}{R_0} := 1 - \theta + a \log \theta = \int_0^\infty \beta_1 y_1(u) y_2(u) \, \mathrm{d}u. \qquad (3.5)$$

It is an elementary consequence of the concavity in $u$ of the function $1 - u + a \log u$, and its gradient, $a - 1$, at $u = 1$, that, for $0 < a < 1$, the function $1 - u + a \log u - C$ has 0, 1, or 2 zeros in $(0, 1]$ according to whether $C$ is greater than, equal to, or less than $1 - a + a \log a$, and that when the last condition holds and $C \geq 0$, the two zeros are located within the intervals $(0, a)$ and $(a, 1)$, respectively. Esteva and Vargas's (1998) analysis shows that starting at any time $t'$ with $X_1(t') > a X_1(0)$ and $Y_1(t') > 0$ leads, for some $t''$, to some $X_1(t'') > 0$ and $Y_1(t'') > 0$, and identifies $\theta$ as lying in $(0, a)$.

When the last term in (3.3) is negligible, $\theta$ is approximately the smaller positive root of

$$R_0(1 - u) + \log u = 0,$$

i.e. $\tilde{\theta}(R_0)$ (see below, above (4.1)); in other words, the fraction of susceptibles surviving is little different from the fraction of susceptibles surviving a standard SIR epidemic with the same initial conditions and parameter $R_0$ as in (3.4). No matter, when $C > 0$, we can conclude that

$$\tilde{\theta}(R_0) < \theta < \frac{1}{R_0}. \qquad (3.6)$$

## 4. Total size in the general SIR epidemic model

For the deterministic general epidemic model in a homogeneously mixing population (see, e.g. Daley and Gani (1999, Chapter 2)), it is known that, at the conclusion of a major outbreak of a disease started from the immigration of an infectious case into a community of $\tilde{N}$ susceptibles, the proportion $\tilde{\theta} := \tilde{\theta}(r)$ of susceptibles surviving is approximately equal to the smaller positive root of the equation

$$\tilde{\theta} = \exp\left[-\frac{1 - \tilde{\theta}}{r}\right], \quad \text{or, equivalently, when } \tilde{\theta} < 1, \quad r = \frac{-\log \tilde{\theta}}{1 - \tilde{\theta}}, \qquad (4.1)$$

where $r$ is the mean number of new infectives produced by a single infectious case introduced into a population in which all other individuals are susceptible. Such a quantity $r$ is often called the basic reproduction ratio $R_0$, and it is simple to check that $\tilde{\theta} < 1$ if and only if $R_0 > 1$ (e.g. use the concavity argument below (3.5)). Formulating the dynamics of this general model via a transmission contact rate $\gamma$ per infectious case and a recovery or removal rate $\beta$ per infectious case, as when there are $X$ susceptibles and $Y$ infectives in a population of total size $\tilde{N}$, we write

$$\dot{X} = -\gamma \frac{X}{\tilde{N}} Y, \qquad \dot{Y} = \gamma \frac{X}{\tilde{N}} Y - \beta Y. \qquad (4.2)$$

Then, under the conditions for interpreting $R_0$ as above, the mean rate at which an infectious case produces new infectives is initially $\gamma X(0)/\tilde{N}$; when these contacts occur with negligible change in $X(\cdot)/\tilde{N}$ for a mean duration $1/\beta$, $r = R_0 = \gamma X(0)/(\beta \tilde{N})$, and $R_0 \approx \gamma/\beta$ when $X(0) \approx \tilde{N}$.

For an epidemic in this SIR model with $R_0 > 1$, there is a critical time $t_{\mathrm{crit}}$ when the number of susceptibles $X(t_{\mathrm{crit}})$ reaches the critical size $X(0)/R_0$ and the number of infectives peaks

at $Y(t_{\mathrm{crit}})$. Then $Z(t_{\mathrm{crit}})/Z(\infty)$ is the proportion of those ultimately infected who are infected by time $t_{\mathrm{crit}}$. By the definition of $t_{\mathrm{crit}}$ and using $\dot{X}$ in (4.2),

$$
\begin{aligned}
Z(t_{\mathrm{crit}}) &= \beta_1 \int_0^{t_{\mathrm{crit}}} Y(u)\, \mathrm{d}u \\
&= -\beta_1 \int_{t=0}^{t=t_{\mathrm{crit}}} \frac{\tilde{N}}{\gamma X}\, \mathrm{d}X \\
&= \frac{\beta_1 \tilde{N}}{\gamma} \int_{1/R_0}^{1} \frac{\mathrm{d}x}{x} \\
&= \frac{\beta_1 \tilde{N}}{\gamma} \log R_0 \\
&= \frac{X(0)}{R_0} \log R_0.
\end{aligned}
$$

Since $Z(\infty) = (1 - \tilde{\theta})X(0)$, with $\tilde{\theta} := \tilde{\theta}(R_0)$ as above (4.1), we have

$$
\frac{Z(t_{\mathrm{crit}})}{Z(\infty)} = \frac{(\log R_0)/R_0}{1 - \tilde{\theta}} = -\frac{\log R_0}{\log \tilde{\theta}}. \tag{4.3}
$$

This relation shows, by virtue of $\tilde{\theta}$ being determined by $R_0$, that the ratio of the size of the supercritical phase of the epidemic to its total size determines $R_0$. In turn, with $R_0$ and $Z(\infty)$ known, we can deduce the value of $X(0)$, since $\tilde{\theta}$ is then determined, and $X(0) = Z(\infty)/(1-\tilde{\theta})$.

In Table 1 we present the values of the ratio (4.3) for $R_0 \geq 1$: the closer $R_0$ is to 1, the closer this ratio is to its maximum value of 0.5. This table also shows that small decreases in the ratio correspond to large increases in $R_0$. Consequently, given the imprecision that can surround

TABLE 1: Proportions of the size of the epidemic and $R_0$.

| $R_0$ | $Z(t_{\mathrm{crit}})/Z(\infty)$ |
|---|---|
| 1.0 | 0.5000 |
| 1.1 | 0.4919 |
| 1.2 | 0.4843 |
| 1.3 | 0.4771 |
| 1.4 | 0.4703 |
| 1.6 | 0.4575 |
| 1.8 | 0.4458 |
| 2.0 | 0.4350 |
| 2.5 | 0.4106 |
| 3.0 | 0.3893 |
| 3.5 | 0.3705 |
| 4.0 | 0.3536 |
| 5.0 | 0.3241 |
| 6.0 | 0.2994 |
| 8.0 | 0.2586 |
| 10.0 | 0.2272 |
| 15.0 | 0.1804 |
| 20.0 | 0.1495 |

reported disease incidence data, particularly (as with DF) when the rate of asymptomatic cases can exceed 90%, the use of the formulae just given to determine $R_0$ or $X(0)$ cannot be expected to yield very precise estimates.

## 5. Total size of an epidemic of a vector-borne disease

We return to the setting where a human disease is spread by an insect vector rather than by direct human contact. Typically, as around (2.2), the duration of a human lifetime is many magnitudes larger than an insect lifetime, the infectious duration of humans may be of the order of magnitude of an insect's lifetime, and the duration of the epidemic is about the length of several mean insect lifetimes. Then we can approximate the dynamics of an epidemic by regarding the human population as of fixed size $N$ for the duration of the epidemic. (In a companion paper in preparation, we consider how this may help our understanding of annual epidemic outbreaks in Singapore.) For (2.1) and (2.2), there are three independent relations:

$$\dot{X}_1 = -\gamma_1 \frac{X_1}{N} Y_2, \tag{5.1a}$$

$$\dot{Y}_1 = \gamma_1 \frac{X_1}{N} Y_2 - \beta_1 Y_1, \tag{5.1b}$$

$$\dot{Y}_2 = \gamma_2 \frac{Y_1}{N} X_2 - \mu_2 Y_2. \tag{5.1c}$$

They are a particular case of a more general model in Esteva and Vargas (1998) in which population sizes are not constant but only stationary.

We can compute the basic reproduction ratio $R_0$ (see, e.g. Daley and Gani (1999, p. 87)) as the product of the mean number of infected mosquitoes produced by a single infectious human introduced into a population of susceptible humans and susceptible mosquitoes (this equals $\gamma_1(1/N)X_2(0)(1/\beta_1) =: r_1$ say), and, in a similar population, the mean number of infectious humans produced by a single infectious mosquito (this equals $\gamma_2(X_1(0)/N)(1/\mu_2) =: r_2$ say), namely,

$$R_0 = r_1 r_2 = \frac{\gamma_1 X_2(0)}{\beta_1 N} \frac{\gamma_2 X_1(0)}{\mu_2 N} = \frac{\gamma_1 \gamma_2 X_1(0) X_2(0)}{\beta_1 \mu_2 N^2} = \frac{\gamma_1 \gamma_2 X_1(0) N_2}{\beta_1 \mu_2 N^2}, \tag{5.2}$$

as in (3.4). Observe that $R_0$ increases as the insect population size $N_2$ (or the capped rate $A$) and initial number of susceptible humans $X_1(0)$ increase; furthermore, $R_0$ would increase with $\gamma_1$ and $\gamma_2$ as a result of increased population density, but this may be offset by a decrease with increased $N$. The other parameters, $\beta_1$ and $\mu_2$, are biological and not dependent on the sociological organization of the populations, except if, for example, infectious humans are isolated from contact with any mosquitoes (this would correspond to increasing $\beta_1$).

Now $Z_1(t) = \beta_1 \int_0^t Y_1(u) \, du$. Equivalently, $z_1(t) = \beta_1 \int_0^t y_1(u) \, du$ after the normalizations

$$(X_1, Y_1, Z_1)(t) = (x_1, y_1, z_1)(t) X_1(0) \quad \text{and} \quad (X_2, Y_2)(t) = (x_2, y_2)(t) N_2.$$

Equations (5.1) can be integrated and manipulated as follows:

$$\log\left(\frac{X_1(0)}{X_1(t)}\right) = \log\left(\frac{1}{x_1(t)}\right) = \frac{\gamma_1}{N} \int_0^t Y_2(u) \, du = \frac{\gamma_1 N_2}{N} \int_0^t y_2(u) \, du, \tag{5.3a}$$

$$y_2(t) = \frac{Y_2(t)}{N_2}$$

$$= \frac{\gamma_2 X_1(0)}{N} \int_0^t y_1(u)\,du - \mu_2 \int_0^t y_2(u)\,du - \frac{\gamma_2 X_1(0)}{N} \int_0^t y_2(u)y_1(u)\,du$$

$$= \frac{\gamma_2 X_1(0)}{\beta_1 N} z_1(t) - \frac{\mu_2 N}{\gamma_1 N_2} \log\left(\frac{1}{x_1(t)}\right) - \frac{\gamma_2 X_1(0)}{\beta_1 N} y_2(\vartheta_{2t})z_1(t) \tag{5.3b}$$

$$= \frac{r_2}{\kappa} z_1(t) - \left[\frac{1}{r_1\kappa} + \frac{r_2}{r_1\kappa} y_1(\vartheta_{1t})\right] \log\left(\frac{1}{x_1(t)}\right). \tag{5.3c}$$

Here $\kappa = \beta_1/\mu_2$ and in the last terms of (5.3b) and (5.3c) we have used the integral mean value theorem so $\vartheta_{jt}$ lies in $(0, t)$ for $j = 1, 2$. In the limit $t \to \infty$, and using $\theta$ again as below (3.2), so

$$\theta = x_1(\infty) = 1 - \frac{Z_1(\infty)}{X_1(0)}, \tag{5.4}$$

we have, for some finite positive $\vartheta_j > 0$ and $j = 1, 2$,

$$0 = \frac{\gamma_2}{\beta_1}(1-\theta)[1 - y_2(\vartheta_2)] + \frac{\log\theta}{r_1} = \frac{\gamma_2}{\beta_1}(1-\theta) + \frac{1}{r_1}\left[1 + \frac{\gamma_2}{\mu_2} y_1(\vartheta_1)\right]\log\theta$$

($\vartheta_j > 0$, else we should have $y_j(\vartheta_j) = 0$ for such $j$, but $\int_0^\infty y_1(u)y_2(u)\,du > 0$).

We summarize our discussion to date in a formal statement which is based on the exact equation for $\theta$, namely,

$$1 - \theta + \frac{\log\theta}{R_0} = \beta_1 \int_0^\infty y_1(u)y_2(u)\,du. \tag{5.5}$$

**Proposition 5.1.** *A vector-borne epidemic in a population as in Section 2 and with dynamics described by the differential equations (2.1) and (2.2) behaves as a major outbreak when $R_0$ in (5.2) is greater than 1. The total size $(1-\theta)X_1(0)$ of such an outbreak has $\tilde{\theta}(R_0) < \theta < 1/R_0$ and satisfies*

$$R_0 := r_1 r_2 := \frac{\gamma_1 N_2}{\beta_1 N} \frac{\gamma_2 X_1(0)}{\mu_2 N} = \frac{-\log\theta}{(1-\theta)[1 - y_2(\vartheta_2)]} = \frac{-\log\theta[1 + (\gamma_2/\mu_2)y_1(\vartheta_1)]}{1-\theta}, \tag{5.6}$$

*where $y_1(\vartheta_1)$ and $y_2(\vartheta_2)$ are intermediate positive values of $Y_1(\cdot)/X_1(0)$ and $Y_2(\cdot)/N_2$, respectively.*

An equivalent version of (5.6) can be phrased in terms of $R_1 = \gamma_1\gamma_2 N_2/(\beta_1\mu_2 N^2)$ as

$$R_1 = \frac{-\log\theta}{X_1(0)(1-\theta)[1 - y_2(\vartheta_2)]} = \frac{-[1 + (\gamma_2/\mu_2)y_1(\vartheta_1)]\log\theta}{X_1(0)(1-\theta)},$$

with $y_1(\vartheta_1)$ and $y_2(\vartheta_2)$ as before. This parameter $R_1$ is the same for any initial number of susceptibles $X_1(0)$ for the same sizes $N$ and $N_2$ of the human and vector populations, respectively, being dependent on only the mosquito contact-and-transmission rate(s) $\gamma_j$ and death rate $\mu_2$, and the human infectious recovery (or 'removal') rate $\beta_1$.

## 6. Further analysis of the vector-borne epidemic differential equations

We return to the differential equation system (5.1a)–(5.1c), deriving upper and lower bounds on the proportion $\theta$ in (5.5), and upper bounds on the quantities $y_j(\vartheta_j)$ ($j = 1, 2$) introduced into (5.4) via the integral mean value theorem. We illustrate the bounds together with values of $\theta$ in Table 2 below, and in Section 7 we indicate that our approach also gives information on the model extended to incorporate two or more categories of humans.

In the models we have considered, it is basically irrelevant whether we use time or the number (or proportion) of susceptibles as the argument of the functions $Y_1(\cdot)$ and $Z_1(\cdot)$ because the mapping $X_1 \colon t \in [0, \infty) \mapsto \{X_1(0) > x > X_1(\infty)\}$ is one-to-one; let $t(X_1)$ denote the functional inverse of $X_1(t)$. It is convenient to replace the argument of $Y_1$ etc. by the standardized susceptible-like variable $x := X_1(t)/X_1(0)$, and to write $y_1(x) = Y_1(t[x X_1(0)])/N_1$ and $y_2(x) = Y_2(t[x X_1(0)])/N_2$. We can now rewrite (5.1a)–(5.1c) as just two differential equations with reference to $X_1$ as the independent variable; in terms of our standardized variables, $r_1$ and $r_2$ in (5.2) and $\kappa$ in (5.3c), they read

$$\frac{dy_1}{dx} = -1 + \frac{1}{r_1} \frac{y_1}{y_2} \frac{1}{x}, \tag{6.1}$$

$$\frac{dy_2}{dx} = -\frac{r_2}{r_1 \kappa} \frac{1 - y_2}{x} \frac{y_1}{y_2} + \frac{1}{r_1 \kappa} \frac{1}{x}.$$

The terms here involving the fraction $y_1/y_2$ are awkward algebraically. We therefore consider the linear combination of these two differential equations, $d(y_1 + cy_2)/dx$, for some constant $c$: we obtain

$$\frac{d(y_1 + cy_2)}{dx} = -1 + \left( \frac{1}{r_1} - c\frac{r_2}{r_1 \kappa} \right) \frac{y_1}{y_2} \frac{1}{x} + c\frac{1}{r_1 \kappa} \frac{1}{x} + c\frac{r_2}{r_1 \kappa} \frac{y_1}{x}.$$

By choosing $c = \kappa/r_2$, the coefficient of $y_1/y_2$ here vanishes, and we obtain

$$\frac{d(y_1 + cy_2)}{dx} = -1 + \frac{1}{r_1 r_2} \frac{1}{x} + \frac{1}{r_1} \frac{y_1}{x} = -1 + \left( \frac{1}{R_0} + \frac{1}{r_1} y_1 \right) \frac{1}{x},$$

where $R_0$ is as defined in (5.2). For as long as $Y_1$ is positive, so too is $Y_2$ (this assertion can be justified much as in Esteva and Vargas (1998)). So, subject to this condition, we have

$$-1 + \frac{1/R_0}{x} \leq \frac{d(y_1 + cy_2)}{dx} \leq -1 + \frac{1/R_0}{x} + \frac{1}{r_1} \frac{y_1 + cy_2}{x}. \tag{6.2}$$

At $x = 1-$, all three expressions here coincide and equal $-(1 - 1/R_0) = -(1 - a)$, so, for sufficiently small $\varepsilon > 0$, each expression has a negative integral over $(1 - \varepsilon, 1)$. Writing $Y(\cdot) = y_1(\cdot) + cy_2(\cdot)$, the integral on $(u, 1)$ of the central expression equals $Y(1) - Y(u) = -Y(u)$, which vanishes at $u = \theta$. Consequently, the smallest nonvanishing intervals $(\theta', 1)$ and $(\theta'', 1)$ over which the integrals of the left- and right-hand sides equal 0, provide the bounding interval $\theta' < \theta < \theta''$ for $\theta$.

The discussion of the standard SIR model (Section 4) shows that, when $R_0 > 1$, there is a root $\tilde{\theta}(R_0)$ within the interval $(0, 1/R_0)$ of

$$R_0(1 - u) + \log u = 0, \tag{6.3}$$

so $\theta' = \tilde{\theta}(R_0)$ and $X_1(\infty) > \tilde{\theta} X_1(0)$. Using (3.6), we have proved the first part of the following.

**Proposition 6.1.** *The fraction $\theta$ of susceptibles surviving a major outbreak of an epidemic as in Proposition 5.1 satisfies*

$$\tilde{\theta}(R_0) < \theta < \check{\theta} < \frac{1}{R_0}, \tag{6.4}$$

*where $\tilde{\theta}(R_0)$ is the smaller positive root of (6.3) and $\check{\theta}$ is the smaller positive zero of the equation*

$$0 = \begin{cases} -\dfrac{1}{b-1}\left(\dfrac{1}{u^{b-1}}-1\right)+\dfrac{1}{bR_0}\left(\dfrac{1}{u^b}-1\right), & b \neq 1, \\[4mm] \log u + \dfrac{1-u}{R_0 u}, & b = 1, \end{cases} \tag{6.5}$$

*where $b = 1/r_1$, with*

$$\check{\theta} = \begin{cases} \tilde{\theta} + \dfrac{bR_0\tilde{\theta}(1-\tilde{\theta})[R_0(1+\tilde{\theta})/2 - 1]}{1 - R_0\tilde{\theta}} + O(b^2) & \text{for } b \to 0, \\[4mm] \dfrac{1-1/b}{R_0} + O(e^{b\log\check{\theta}}) & \text{for } b \to \infty. \end{cases} \tag{6.6}$$

*Proof.* It remains to prove the assertions about $\check{\theta}$ in (6.5) and (6.6). Using $Y(x) := y_1(x) + cy_2(x)$, rewrite (6.2) as

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = -1 + \frac{a + bY(x)}{x} - \frac{bcy_2(x)}{x}, \qquad 0 < a = \frac{1}{R_0} < 1, \ b = \frac{1}{r_1} > 0, \tag{6.7}$$

for as long as $\theta < x < 1$, which is the interval where $y_j(x)$, $j = 1, 2$, are both positive. Introduce the function $V(x)$ on $0 < x \leq 1$ as the solution of the differential equation

$$\frac{\mathrm{d}V}{\mathrm{d}x} = -1 + \frac{a + bV(x)}{x}, \qquad V(1) = 0; \tag{6.8}$$

as an aside, since $Y(1) = V(1) = 0$, $V'(1-) = Y'(1-) = -(1-a)$. By taking differences of the respective sides of (6.7) and (6.8) we deduce that

$$\frac{\mathrm{d}(Y-V)}{\mathrm{d}x} - \frac{b(Y-V)(x)}{x} = -\frac{bcy_2(x)}{x}, \quad \text{i.e.} \quad \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{Y-V}{x^b}\right) = -\frac{bcy_2(x)}{x^{b+1}} \tag{6.9}$$

on making use of an integrating factor. (For the moment, we ignore the case in which $b = 1$; it is treated around (6.10) below.) Integrating over the interval $(z, 1)$ for any $z \in (0, 1)$ and using $Y(1) = V(1) = 0$, we have

$$\frac{Y(z) - V(z)}{z^b} = \int_z^1 bcu^{-1}y_2(u)\,\mathrm{d}u > 0,$$

implying that $Y(z) > V(z)$ for all such $z$. But $Y(z) \geq 0$ for $\theta \leq z \leq 1$, with equality only at the two endpoints of the interval. Consequently, if $V(x) = 0$ for any $x \in (\theta, 1)$, $x = \check{\theta}$ say, we must have $\theta < \check{\theta}$. We use $\check{\theta}$ in place of $\theta''$ defined above (6.3).

It is readily checked from (6.8) that $x^{-b}V(x)$ is concave on $(0, 1]$, so $V(\cdot)$ can have at most two zeros on that interval: there is one at 1 by construction, and the other lies in $(0, a)$. We must therefore have $\theta < \check{\theta} < a$, as asserted in (6.4).

Rewrite (6.8) so as to incorporate an integrating factor, much as in (6.9). Integration yields

$$\frac{V(x)}{x^b}\bigg|_{x=u}^{x=1} = \begin{cases} -\dfrac{1}{b-1}\left(\dfrac{1}{u^{b-1}}-1\right)+\dfrac{a}{b}\left(\dfrac{1}{u^b}-1\right), & b>0,\ b\neq 1, \\[3mm] \log u + \dfrac{a(1-u)}{u}, & b=1; \end{cases} \quad (6.10)$$

hence (6.5). Write the general case (i.e. $b\neq 1$) of (6.5) for $\check{\theta}$ in the form

$$R_0 u = \left(1-\frac{1}{b}\right)+\left(R_0-\left(1-\frac{1}{b}\right)\right)^b = \left(1-\frac{1}{b}\right)+\left(R_0-\left(1-\frac{1}{b}\right)\right)e^{b\log u}. \quad (6.11)$$

Expanding the latter form of (6.11) as a power series in $b$ for small $b>0$ yields

$$0 = R_0(1-u)+\log u + b(\log u)\left(R_0-1+\tfrac{1}{2}\log u\right)+O(b^2)=:g_b(u) \quad \text{say,}$$

for which $g_0(u) = R_0(1-u)+\log u$, $g_0(\tilde{\theta})=0$. Otherwise, using a Newton–Raphson technique, the root $\check{\theta}(R_0,b)$ for $b$ near 0 is given by $\check{\theta}\approx\tilde{\theta}-g_b(\tilde{\theta})/g'_b(\tilde{\theta})$, where

$$g'_b(u) = -R_0+\frac{1}{u}+\frac{b}{u}(R_0-1+\log u)+O(b^2)$$

(here and elsewhere, $O(\cdot)$ need not be the same function). Now set $\log\tilde{\theta}=-R_0(1-\tilde{\theta})$ and simplify; this gives the first case of (6.6).

No matter what the value of $b<\infty$, $\check{\theta}<1/R_0<1$ so, for large $b$ and $u=\check{\theta}$, $u^b=e^{b\log u}=o(1/b)$. Hence, the first part of (6.11) implies the relation for $\check{\theta}$ for large $b$ as asserted in (6.6). This completes the proof.

Table 2 illustrates how $\check{\theta}(R_0,r_1)$ varies with $R_0$ on $(1,\infty)$ and $r_1$ on $(0,\infty)$, together with $\tilde{\theta}(R_0)=\check{\theta}(R_0,\infty)$. We have also computed $\theta$ for the same values of $r_1$ and $r_2=R_0/r_1$ for the biologically ball-park cases of $\kappa=\beta_1/\mu_2=1,2$. DD and RS did this, independently, using MATLAB$^{\circledR}$ and MATHEMATICA$^{\circledR}$ differential equation solvers.

The tabulated results show that, for the model as we have discussed it, the proportion of initial infectives affected in a major outbreak ranges from 'small' (say, less than 10%) to 'most' (say, more than 80%) with $R_0$ changing from 1 to about 3. In other words, when the level of susceptibles has come up to the critical value, a three-fold increase in these numbers is enough to result in any dengue epidemic affecting the vast majority of such individuals.

The product construction leading to $R_0 = r_1 r_2$ in (5.2) reflects the two infection (*quasi branching*) processes in the model. The extreme values of $b = 1/r_1$, for given $R_0 > 1$, reflect situations where one or other of these processes is pre-eminent in making the process supercritical as a whole.

The lower bound in (6.2) has an explicit integral, and this shows that

$$1-u+\frac{\log u}{R_0} > y_1(u)+cy_2(u), \qquad \theta<u<1.$$

Setting $a=1/R_0$, the supremum of the left-hand side for the range of $u$ as shown occurs for $u=a$ and equals $1-a+a\log a$. Appealing to nonnegativity, this supremum is an upper bound on both $y_1(\vartheta_1)$ and $cy_2(\vartheta_2)$ in Proposition 5.1. The bound equals

$$(1-a)^2\left(1-\tfrac{1}{2}a\right)-a(1-a)^3\left(\tfrac{1}{3}+\tfrac{1}{4}(1-a)+\tfrac{1}{5}(1-a)^2+\cdots\right).$$

TABLE 2: Bounds $\check{\theta}(R_0, r_1)$ and proportions $\theta(R_0, r_1, \kappa)$ of initial susceptibles surviving major outbreak.

| $r_1$ | $R_0$ | $\check{\theta}(R_0, r_1)$ | $\theta(R_0, r_1, \kappa)$ | |
|---|---|---|---|---|
| | | | $\kappa = 1$ | $\kappa = 2$ |
| 0.0 | 1.10 | 0.909 09 | — | — |
| | 1.25 | 0.800 00 | — | — |
| | 1.50 | 0.666 67 | — | — |
| | 2.00 | 0.500 00 | — | — |
| | 3.00 | 0.333 33 | — | — |
| | 5.00 | 0.200 00 | — | — |
| 0.1 | 1.10 | 0.857 07 | 0.8438 | 0.8379 |
| | 1.25 | 0.732 44 | 0.7018 | 0.6848 |
| | 1.50 | 0.602 52 | 0.5622 | 0.5351 |
| | 2.00 | 0.450 19 | 0.4134 | 0.3837 |
| | 3.00 | 0.300 00 | 0.2745 | 0.2522 |
| | 5.00 | 0.180 00 | 0.1647 | 0.1511 |
| 0.2 | 1.10 | 0.844 23 | 0.8356 | 0.8309 |
| | 1.25 | 0.700 89 | 0.6741 | 0.6616 |
| | 1.50 | 0.558 75 | 0.5156 | 0.4919 |
| | 2.00 | 0.406 67 | 0.3583 | 0.3263 |
| | 3.00 | 0.267 67 | 0.2293 | 0.1998 |
| | 5.00 | 0.160 09 | 0.1361 | 0.1164 |
| 0.5 | 1.10 | 0.833 33 | 0.8277 | 0.8257 |
| | 1.25 | 0.666 67 | 0.6498 | 0.6431 |
| | 1.50 | 0.500 00 | 0.4672 | 0.4527 |
| | 2.00 | 0.333 33 | 0.2882 | 0.2652 |
| | 3.00 | 0.200 00 | 0.1574 | 0.1320 |
| | 5.00 | 0.111 11 | 0.0818 | 0.0619 |
| 1.0 | 1.10 | 0.828 87 | 0.8248 | 0.8231 |
| | 1.25 | 0.649 96 | 0.6395 | 0.6356 |
| | 1.50 | 0.466 41 | 0.4446 | 0.4359 |
| | 2.00 | 0.284 67 | 0.2514 | 0.2369 |
| | 3.00 | 0.149 00 | 0.1150 | 0.0984 |
| | 5.00 | 0.069 92 | 0.0462 | 0.0331 |
| $R_0$ | 1.10 | 0.828 44 | 0.8245 | 0.8228 |
| | 1.25 | 0.646 13 | 0.6373 | 0.6340 |
| | 1.50 | 0.452 26 | 0.4360 | 0.4298 |
| | 2.00 | 0.250 00 | 0.2289 | 0.2207 |
| | 3.00 | 0.095 71 | 0.0792 | 0.0728 |
| | 5.00 | 0.019 57 | 0.0133 | 0.0111 |
| 5.0 | 1.10 | 0.824 92 | 0.8194 | 0.8152 |
| | 1.25 | 0.633 37 | 0.6293 | 0.6271 |
| | 1.50 | 0.428 86 | 0.4225 | 0.4200 |
| | 2.00 | 0.223 81 | 0.2138 | 0.2101 |
| | 3.00 | 0.082 16 | 0.0714 | 0.0675 |
| | 5.00 | 0.019 57 | 0.0133 | 0.0111 |
| $\infty$ ($\tilde{\theta}(R_0)$) | 1.10 | 0.823 87 | — | — |
| | 1.25 | 0.628 63 | — | — |
| | 1.50 | 0.417 19 | — | — |
| | 2.00 | 0.203 19 | — | — |
| | 3.00 | 0.059 52 | — | — |
| | 5.00 | 0.006 98 | — | — |

## 7. A stratified population model approach

Seroprevalence surveys for DF antibodies in Singapore suggest that the contact-and-transmission rates there between mosquitoes and humans are rather lower post-1970 for children than for adults. We therefore investigate what modelling might yield in terms of total sizes of annual cases within two age groups with different contact rates. Let $X_{1j}$ and $Y_{1j}$ denote the numbers of susceptibles and infectives in those under and over age 15 (say) ($j = 1, 2$, respectively). Let $\gamma_{1j}$ and $\gamma_{2j}$ denote the respective rates for the contact and transmission of infection from mosquito to human and human to mosquito for group $j$.

We can construct differential equations for this model much as for (2.1) and (2.2):

$$\dot{X}_{1j} = -\gamma_{1j}\frac{X_{1j}}{N}Y_2, \qquad \dot{Y}_{1j} = \gamma_{1j}\frac{X_{1j}}{N}Y_2 - \beta_1 Y_{1j}, \qquad \dot{Z}_{1j} = \beta_1 Y_{1j}, \qquad (7.1)$$

for $j = 1, 2$, and

$$\dot{Y}_2 = \frac{\gamma_{21}Y_{11} + \gamma_{22}Y_{12}}{N}X_2 - \mu_2 Y_2. \qquad (7.2)$$

To compute $R_0$, we find that the mean number of group $j$ infectives produced by a single infectious mosquito in a population consisting of all susceptibles equals $\gamma_{1j}X_{1j}(0)/(\mu_2 N)$, and the mean number of infected mosquitoes produced by a single group $j$ infective in a similar population equals $\gamma_{2j}X_2(0)/(\beta_1 N)$, so

$$R_0 = \frac{\gamma_{11}\gamma_{21}X_{11}(0) + \gamma_{12}\gamma_{22}X_{12}(0)}{\beta_1 \mu_2 N^2}N_2. \qquad (7.3)$$

The differential equations for $\dot{X}_{1j}$ in (7.1) show that these two variables are functionally dependent in an explicit manner, since

$$\log\left(\frac{X_{1j}(0)}{X_{1j}(t)}\right) = \frac{\gamma_{1j}}{N}\int_0^t Y_2(u)\,du. \qquad (7.4)$$

Also, since $Z_{1j}(t) = \beta_1 \int_0^t Y_{1j}(u)\,du$, integration of (7.2) yields

$$Y_2(t) = \frac{N_2}{\beta_1 N}(\gamma_{21}Z_{11}(t) + \gamma_{22}Z_{12}(t)) - \frac{\mu_2 N}{\gamma_{12}}\log\left(\frac{X_{12}(0)}{X_{12}(t)}\right)$$

$$- \frac{1}{N}\int_0^t (\gamma_{21}Y_{11}(u) + \gamma_{22}Y_{12}(u))Y_2(u)\,du. \qquad (7.5)$$

Equation (7.4) shows that the ultimate proportions $\theta_{1j} := X_{1j}(\infty)/X_{1j}(0)$ are related by $\theta_{11} = \theta_{12}^{\gamma_{11}/\gamma_{12}}$, and then (7.5) shows that $\theta_{12}$ satisfies a relation analogous to (3.4), from which we can deduce our final result by a similar argument to the one following (3.4).

**Proposition 7.1.** *The fractions $\theta_{1j}$ of susceptibles surviving a major epidemic of a disease, similar to Proposition 5.1 but for a stratified population of two groups as above, satisfy $\theta_{11} = \theta_{12}^{\gamma_{11}/\gamma_{12}}$, and $\theta_{12}$ lies in $(\theta', \theta'')$, where $\theta'$ is the smaller positive root of $f(u) = 0$ and*

$$f(u) := \gamma_{12}\gamma_{21}X_{11}(0)(1 - u^{\gamma_{11}/\gamma_{12}}) + \gamma_{12}\gamma_{22}X_{12}(0)(1 - u) + \frac{\beta_1 \mu_2 N^2}{N_2}\log u, \qquad (7.6)$$

*and $\theta''$ is the root in $(0, 1)$ of $f'(u) = 0$; we can write this equation as*

$$\gamma_{11}\gamma_{21}X_{11}(0)u^{\gamma_{11}/\gamma_{12}} + \gamma_{12}\gamma_{22}X_{12}(0)u = \frac{\beta_1 \mu_2 N^2}{N_2}. \qquad (7.7)$$

We assert that refinements along the lines of Section 6 are also possible by rewriting the differential equations in (7.1) and (7.2) much as in (6.1). This leads to three differential equations of which the linear function $c_1 Y'_{11} + c_2 Y'_{12} + Y'_2$, where the prime signifies a derivative with respect to $\xi := X_{12}$, yields an inequality similar to (6.2) and, hence, bounds on the ultimate proportions $\theta_{1j}$.

The analysis of this section extends immediately to groups $j = 1, \ldots, J$ for a finite positive integer $J$ and positive rates $\gamma_{1j}$ and $\gamma_{2j}$, with, e.g. $f$ in (7.6) replaced by

$$f(u) = \gamma_{1J} \sum_{j=1}^{J} \gamma_{2j} X_{1j}(0)(1 - u^{\gamma_{1j}/\gamma_{1J}}) + \frac{\beta_1 \mu_2 N^2}{N_2} \log u,$$

while (7.7) becomes

$$\sum_{j=1}^{J} \gamma_{1j} \gamma_{2j} X_{1j}(0) u^{\gamma_{1j}/\gamma_{1J}} = \frac{\beta_1 \mu_2 N^2}{N_2}. \tag{7.8}$$

The analogue of $R_0$ in (7.3) equals the left-hand side of (7.8) evaluated at $u = 1$ and divided by the right-hand side.

## Acknowledgements

## References

DALEY, D. J. (2010). Comment on the paper by Gani. *Austral. N. Z. J. Statist.* **52,** 330–334.

DALEY, D. J. AND GANI, J. (1999). *Epidemic Modelling: An Introduction* (Camb. Stud. Math. Biol. **15**). Cambridge University Press.

DIETZ, K. (1974). Transmission and control of arbovirus diseases. In *Epidemiology*, eds D. Ludwig and K. L. Cooke, Society for Industrial and Applied Mathematics, Philadelphia, PA, pp. 104–121.

ESTEVA, L. AND VARGAS, C. (1998). Analysis of a dengue disease transmission model. *Math. Biosci.* **150,** 131–151.

GANI, J. (2010). Modelling epidemic diseases. *Austral. N. Z. J. Statist.* **52,** 321–329.

GUBLER, D. J. AND KUNO, G. (eds) (1997). *Dengue and Dengue Hemorrhagic Fever*. CAB International, Wallingford, Oxon.

OOI, E.-E., GOH, K.-T. AND GUBLER, D. J. (2006). Dengue prevention and 35 years of vector control in Singapore. *Emerging Infectious Diseases* **12**, 887–893.

DARYL J. DALEY, *The Australian National University and The University of Melbourne*

Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia.
Email address: daryl.daley@anu.edu.au

RANDALL J. SWIFT, *California State Polytechnic University, Pomona*

Department of Mathematics, California State Polytechnic University, Pomona, CA 91768, USA.
Email address: rjswift@csupomona.edu