Supporting information 1. Supplementary model description

Supplement to:

Impact of age-specific immunity on the timing and burden of the next Zika virus outbreak

Michel J. Counotte, Christian L. Althaus, Nicola Low and Julien Riou

Contents

S1.1 Model workflow
S1.2 Comparison of SIR model with SEIR model and the Pandey model.
S1.3 Prior predictive check
S1.4 Comparison of “square root normal”, “negative binomial” and “Poisson” approaches
S1.5 $R_0$ calculation
S1.6 Loss of immunity scenarios
S1.7 ABM algorithm
S1.8 The number of infections introduced does influence the probability of an outbreak, but not the attack rate of successful outbreaks
S1.1 Model workflow

Figure 1 provides a workflow of the parameterization using an ODE model and subsection predictions using an ABM model.

Figure 1: Graphic description of the model.

Code available: Github: ZikaProject/SeroProject
S1.2 Comparison of SIR model with SEIR model and the Pandey model.

We chose a simple SIR structure to model the transmission of ZIKV (Fig. 2). Other common choices include SEIR structures, including an incubation period, and a Pandey-type structure explicitly modelling the vector population as implemented in Champagne et al. (2016) [1]. We support this choice by conducting model selection using leave-one-out cross-validation (LOO-CV) [2]. The objective of LOO-CV is to estimate the leave-one-out information criterion (LOOIC), a measure of the pointwise out-of-sample prediction accuracy from a fitted Bayesian model. The estimation of the LOOIC relies on Pareto smoothed importance sampling (PSIS), a procedure for regularizing importance weights.

![Diagram showing SIR, SEIR, and Pandey models]

Figure 2: Description of the three compared models. These diagrams ignore the stratification in two age groups: 0-14 and ≥15.

We compared the fits of the SIR, SEIR and Pandey model to incidence and seroprevalence data from Managua, Nicaragua. The LOOIC for the SEIR model (93.9) was slightly lower than for the SIR model (95.2) but the the estimated pointwise difference in LOOIC (ΔLOOIC) of -1.3 was small compared to its standard error, indicating no evidence in support of a better fit. The ΔLOOIC between the SIR and Pandey models was also small and in favour of the SIR model. Overall, this model selection approach supported our choice of the SIR model.

<table>
<thead>
<tr>
<th>Model</th>
<th>LOOIC (SE)</th>
<th>ΔLOOIC (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIR</td>
<td>95.2 (8.7)</td>
<td>Ref.</td>
</tr>
<tr>
<td>SEIR</td>
<td>93.9 (8.9)</td>
<td>-1.3 (1.6)</td>
</tr>
<tr>
<td>Pandey</td>
<td>99.5 (8.3)</td>
<td>+4.3 (3.6)</td>
</tr>
</tbody>
</table>

S1.3 Prior predictive check

The choice of prior distribution is a crucial aspect of analyses conducted in a Bayesian framework. Prior predictive checks can be used to assess the adequacy of the choice of prior distributions [3]. The principle is to use the model to simulate artificial data from the chosen set of prior distribution. If the chosen set of priors can lead to any dataset that could plausibly be observed, then the priors can be qualified as “non-informative”. Figure 3 shows that this is indeed the case here, as our choice of priors for $\beta_1$, $\beta_2$, $\gamma$, $\rho$...
and \( I(0) \) lead to a wide variety of possible epidemic data, from 0 to 4,000 cases reported weekly and from 0 to 100% post-epidemic seroprevalence.

![Graph A: Weekly number of reported cases of Zika virus infection in Managua, Nicaragua (grey area shows the 95% range in the artificial data simulated from the prior distributions, red dots show actual data).
Graph B: Post-epidemic seroprevalence (grey bar shows the 95% range in the artificial data, red dots are actual data).](image)

Figure 3: Prior predictive check for the model. (A) Weekly number of reported cases of Zika virus infection in Managua, Nicaragua (grey area shows the 95% range in the artificial data simulated from the prior distributions, red dots show actual data). (B) Post-epidemic seroprevalence (grey bar shows the 95% range in the artificial data, red dots are actual data).

**S1.4 Comparison of “square root normal”, “negative binomial” and “Poisson” approaches**

Several approaches can be used to fit an ODE model to incidence data. The objective is to obtain the joint posterior distribution of the parameters \( \beta_1, \beta_2 \) and \( \gamma \) by considering the likelihood of the incidence data \( A \). A straightforward choice for modelling count data is to use a Poisson distribution to link the output of the ODE system at time \( t \) \( D_t \) to weekly incidence data \( A_t \):

\[
\Pr(A|\beta_1, \beta_2, \gamma) = \prod_t \text{Poisson}(A_t|C_t) \quad (1)
\]

A common problem using Poisson distributions is the presence of overdispersion. A direct solution is to use instead a negative binomial distribution, with an additional overdispersion parameter \( \phi \):

\[
\Pr(A|\beta_1, \beta_2, \gamma) = \prod_t \text{Neg-Bin}(A_t|C_t, \phi) \quad (2)
\]

However, the classical negative binomial distribution can struggle when data varies from 0 to large values, as it results in variance estimates that do not scale properly. A solution is to use a modified negative binomial distribution where the overdispersion parameter is scaled by the mean, so that:

\[
\Pr(A|\beta_1, \beta_2, \gamma) = \prod_t \text{Neg-Bin}(A_t|C_t, C_t \times \phi) \quad (3)
\]

An alternative is to use a normal distribution after a square-root transformation aimed at stabilizing the variance, as described in [4]:

\[
\Pr(A|\beta_1, \beta_2, \gamma) = \prod_t \mathcal{N}((\sqrt{A_t}|\sqrt{C_t}, \sigma) \quad (4)
\]

We decided to use this last solution in our model, but provide here a comparison of the model fit (Figure 4) and parameter estimates (Table 2) obtained with the other approaches. We show that all the approaches lead to very similar fits and parameter estimates, although the Poisson approaches leads to narrower credible intervals.
Figure 4: Comparison of fit with Sqrt (A), Poisson (B), and Modified negative binomial (C).

Table 2: Comparison of parameter estimates with Sqrt, negative binomial, and Poisson.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Square root normal</th>
<th>Poisson</th>
<th>Modified negative binomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.17 (0.15–0.20)</td>
<td>0.16 (0.14–0.18)</td>
<td>0.17 (0.15–0.19)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.31 (0.28–0.34)</td>
<td>0.30 (0.27–0.32)</td>
<td>0.31 (0.29–0.33)</td>
</tr>
<tr>
<td>$1/\gamma$</td>
<td>5.24 (4.69–5.86)</td>
<td>5.58 (4.98–6.21)</td>
<td>5.35 (4.85–5.84)</td>
</tr>
</tbody>
</table>

S1.5 $R_0$ calculation

We used the next generation matrix method described by Diekmann et al. (2010) to calculate $R_0$ (eq. 5 - 7) [5]. $\beta_1$ is the transmission rate for the 0–14 age group; $\beta_2$ for the >15 group and $\gamma$ is the common recovery rate.

We start by expressing the model with the infection matrix $F$ and the migration matrix $V$:

\[
F = \begin{pmatrix} \beta_1 & \beta_1 \\ \beta_2 & \beta_2 \end{pmatrix}
\]

(5)

\[
V = \begin{pmatrix} -\gamma & 0 \\ 0 & -\gamma \end{pmatrix}
\]

(6)

$R_0$ is defined as the square root of the largest eigenvalue of $FV^{-1}$:

\[
R_0 = \sqrt{\frac{\beta_1 + \beta_2}{\gamma}}
\]

(7)

S1.6 Loss of immunity scenarios

We explored plausible scenarios of loss of immunity with mean durations of 15, 30, 60, 90, and 150 years (Fig. 5).

Figure 5: Different scenarios considered regarding the loss of immunity.
Here, we provide pseudo code of the ABM (Algorithm 1).

**Algorithm 1 ABM**

1: `procedure` **Initialization**  \(
\triangleright \text{Add initial conditions S/R and sex per n individual}
\)

2: `for` \(n \leftarrow 1, \text{popMax} \) `do`

3: \( R[n] \leftarrow \text{select random 1 or 0 with probability(age[n])} \)

4: \( S[n] \leftarrow 1 - R[n] \)

5: \( I[n] \leftarrow 0 \)

6: \( \text{sex}[n] \leftarrow \text{select random 1 or 0 with probability 0.5} \)

7: `end for`

8: `end procedure`

9: `procedure` **Simulation**  \(
\triangleright \text{Simulation over wkMax weeks}
\)

10: `for` \(wk \leftarrow 1, \text{wkMax} \) `do`

11: `for` \(n \leftarrow 1, \text{popMax} \) `do`

12: `if` \(\text{individual is alive} \) `then`

13: `procedure` **Population dynamics**  \(
\triangleright \text{Pre-outbreak}
\)

14: `Birth, Death, Ageing`

15: `end procedure`

16: `procedure` **Loss of Immunity**  \(
\triangleright \text{Loss of immunity}
\)

17: \([R \rightarrow S]\) \text{ with probability RateToProb(\(\xi\))}

18: `end procedure`

19: `procedure` **Vaccination**  \(
\triangleright \text{Vaccination}
\)

20: \([S \rightarrow R]\) \text{ with probability vaccinationProb, at age}[n]

21: `end procedure`

22: `procedure` **Infection, Recovery**  \(
\triangleright \text{During outbreak}
\)

23: \([S \rightarrow I]\) \text{ with probability RateToProb(\(\beta, \text{age}[n]\))}

24: \([I \rightarrow R]\) \text{ with probability RateToProb(\(\gamma\))}

25: `end procedure`

26: `end if`

27: `end for`

28: `procedure` **Start outbreak**  \(
\triangleright \text{Introduction of infection}
\)

29: `if` \(wk = \text{introductionWk} \) `then`

30: `Change timestep: 7 days to 0.1 days`

31: `Collect summary statistics pre-outbreak`

32: `Introduce introductionN infections`

33: `end if`

34: `end procedure`

35: total number alive  \(
\triangleright \text{Collect summary of week wk:}
\)

36: total number infected

37: `end for`

38: `end procedure`
The number of infections introduced does influence the probability of an outbreak, but not the attack rate of successful outbreaks.

The proportion of outbreaks (1% threshold) after introduction depends on the number of infections introduced; the attack rate of the successful outbreaks does not depend on the number of infections introduced (Fig. 6).

Figure 6: Attack rate over time for the introduction of (A) n=1, (B) n=5, (C) n=10 infections.

References


