Supplementary Information Appendix for
Time-varying, serotype-specific force of infection of dengue virus

Robert C. Reiner, Jr., Steven T. Stoddard, Brett M. Forshey, Aaron A. King, Alicia M. Ellis, Alun L. Lloyd, Kanya C. Long, Claudio Rocha, Stalin Vilcarromero, Helvio Astete, Isabel Bazan, Audrey Lenhart, Gonzalo M. Vazquez-Prokopec, Valerie A. Paz-Soldan, Philip J. McCall, Uriel Kitron, John Elder, Eric Halsey, Amy Morrison, Tadeusz J. Kochel Thomas W. Scott

correspondence to: rcreiner@ucdavis.edu

SECTION S1 Longitudinal cohorts

Seroconversion data used in our analysis were collected from five longitudinal cohort studies carried out between January 1999 and December 2010 (Table S1). When participants left the study, sometimes temporarily, they were replaced by recruiting other residents from the same geographic area. Individual participants provided between 1-13 samples separated typically by 6-9 months. In Figure S1, we plotted the frequency of individuals entering and leaving the study, as indicated by their first and last blood sample, respectively. Enrollment numbers were highest at the initiation of individual cohort studies, specifically January-September 1999, April-May 2004, August-November 2006, November 2007-May 2008 and August-October 2009. Numbers were high throughout the 1999-2005 cohort because sampling and recruitment were staggered temporally (Fig. S1a). Most participants left the studies at the end of each cohort (Fig. S1b), but individuals were lost to follow up at other times for various reasons including: moving away from the study area, opting to drop out of the study, and death. Although each cohort study had distinct objectives, for each we collected blood samples longitudinally and tested them by plaque reduction neutralization test ($PRNT_{70}$) to identify seroconversion to all 4 DENV serotypes.

SECTION S2 Serotype identification of infections and alternative seroconversion identification algorithms

The plaque reduction neutralization test (PRNT) is considered the gold standard for DENV serology. Although specific for each DENV serotype, interpretation of PRNT results is complicated by cross-reactions among DENV-neutralizing antibodies. A novel infection by one serotype will thus cause an increase in titers of neutralizing antibodies to other serotypes. Moreover, an individual infected by a heterologous serotype may have a more robust antibody response to the first than second serotype with which they were infected; i.e., 'original antigenic sin' [1, 2]. Analyzing multiple samples longitudinally is, however, a way to ameliorate uncertainty due to cross-reaction, immunological 'noise', and test error; i.e., sensitivity/specificity [3, 4].

Serotype identification of infections

Serum samples were analyzed for serotype-specific neutralizing antibodies using a plaque reduction neutralization test (PRNT), modified from Morens et al [3, 5, 6]. Test sera were heat inactivated at $56°C$ for 30 minutes prior to dilution. Diluted test sera (0.2 mL) were mixed with 0.2 mL diluted media [Earle's minimal essential medium (E-MEM) with 2% fetal bovine serum (FBS) and antibiotic/antimycotic] containing 40-80 PFU of assay virus and then incubated at $4°C$ for 15 hours. Virus-serum mixture (0.1 mL) was added in triplicate to 0.5 mL

media containing 1.5x105 baby hamster kidney-21 (BHK21) cells and then added to a well of a 24 well tissue culture plate. Plates were incubated at $37°C$ with 5% CO_2 for approximately 3 hrs. Overlay media (0.5 mL of 0.6% carboxymethyl Cellulose, E-MEM w/o phenol Red, 10% FBS, 0.075% $NaHCO_3$ and antibiotic/antimycotic) was added to the adhered cells and incubated at $37°C$ in 5% CO_2 for multiple days (depending on the serotype). Following incubation, the overlay media was removed, and the cells were rinsed (with water) and stained [0.5 mL of 0.1% (w/v) Naphthol Blue Black, 1.36% (w/v) Sodium Acetate, and 6% (v/v) Glacial Acetic Acid] for 30 min. Stain was then removed, the BHK-21 monolayers were washed and air dried, and plaques were counted manually. Between 1999 and 2005, sera were tested at dilutions (after the addition of virus) of 1:60 and 1:120 for DENV-1 and DENV-2, and 1:30 and 1:60 or 1:60 and 1:120 for DENV-3. From 2006 to 2010, sera were tested at dilutions of 1:40, 1:80, 1:160, and 1:640 (after the addition of virus). Results were expressed as the serum dilution (based on a linear fit (1999-2005) or a probit regression fit (2006-2010)) that reduced the number of plaques by 70% (i.e., $PRNT_{70}$) relative to normal human serum at the same dilution. Between 1999 and 2005, for DENV-2, linear regression models were fit to estimate the percent reduction at a cut-off dilution of 1:80, whereas a cutoff dilution of 1:60 was used for DENV-1 and DENV-3. Between 2006 and 2010, titers were based on probit regression fit to the four dilution series (1:40, 1:80, 1:160, and 1:640). To address continuity and comparability between methods used for the two times frames (1999-2005 and 2006-2010), we compared classification from linear regression models of two dilutions with probit models of four dilutions. We focused on samples that fell between the two dilutions and thus would be closer to the cutoff dilution and more difficult to properly classify. We estimate a concordance of 87% between the two dilution and four dilution approaches (84% sensitivity and 95% specificity, using four dilution as a gold standard). Positive and negative control human sera were included with each set of samples analyzed.

Viruses utilized in the assay were amplified in *Ae. albopictus* C6/36 cell culture and frozen in aliquots at $-70°C$ at various points over the course of the study to increase standardization. Test viruses were DENV-1 16007 (DHF case from Thailand, 1964), DENV-2 16681 (DHF case from Thailand, 1964), DENV-3 IQD1728 (DF case from Peru, 2002), and DENV-4 1036 (DF case from Indonesia, 1976). Cut-off dilutions were set at 1:60 for DENV-1, 1:80 for DENV-2, 1:60 for DENV-3, and 1:40 for DENV-4. The cut-offs were selected to balance maximizing sensitivity and specificity, based on results from our laboratory as described in [7, 3, 4, 8]. Seroconversions were based on an increase in reciprocal neutralizing titers from below the cutoff to above the cutoff between paired blood samples.

Alternative seroconversion identification algorithms

Prior to analysis the raw serology data was processed to identify putative DENV infections based on the titer of serotype-specific, neutralizing antibodies. Because DENV serology data is notoriously noisy [9], our approach aimed to reduce uncertainty. Our algorithm was designed to account for the fact that each DENV serotype (and genotype) is a unique virus with its own (biological) PRNT assay and that neutralizing DENV antibodies cross-react on short and long time scales, although the exact nature of these interactions remain poorly understood. In this regard, it is important to note that the dengue season in Iquitos is usually ≤ 6 months, which is consistent with the time frame of temporary, heterologous cross-protection and elevated IgM antibodies; something that would limit multiple, sequential infections with different serotypes in the same transmission season. This, in combination with Iquitos clinic data showing that almost all viruses recovered from dengue patients during a given transmission season were a single serotype, this means that infection with more than one DENV serotype in the same person in a single season was unlikely. Our base algorithm was thus as follows:

1. If a single antibody titer (described in the methods of the main text) exceeded the serotype specific threshold, the results was considered positive. If the result was below the threshold, it was considered negative.

2. Any transiently positive result in a series of samples from the same individual (i.e., if a person presented as positive - negative positive) the positive result was considered to be false and converted to a negative result.

3. If a person appeared to seroconvert to more than one serotype between two serial samples (i.e., they were positive to DENV-1 in the first sample and positive to DENV-2 and DENV-3 in the second sample) that person was removed from the study.

This cleaning process eliminated potentially confounding, false positive results. Because we expect that some results categorized this way were not false, this algorithm was conservative because it retained fewer seroconversions than actually occurred. Nevertheless, we considered a number of alternative algorithms that were both more and less restrictive than the algorithm described above. Generally, with the exception of the most restrictive algorithms, results were comparable to those we present in the main text. Moreover, in all cases we observed seroconversions to multiple serotypes in the same year, which conflicts with Iquitos clinic data. The alternative algorithms, organized from least to most restrictive, focused on the following three key issues.

Multiple serotype seroconversion Our base algorithm eliminated any instances of multiple serotype seroconversion. There were, however, individuals from whom blood samples were separated by long (>10 months) intervals. In these instances the person could have been infected twice in two different transmission seasons, which would have looked like a multiple seroconversion in their serology. We, therefore, relaxed our criterion in two ways: (1) by allowing multiple serotype seroconversions if the serial interval was ≤10 months and (2) by allowing all multiple serotype conversions. The latter was included primarily to assess the sensitivity of our results to the elimination of all individuals who appeared to seroconvert to multiple serotypes in the base algorithm.

PRNT sensitivity/specificity Prior to their introduction into Iquitos, we detected almost no individuals with antibodies to either DENV-3 or DENV-4. These assays are thus highly specific, although their sensitivity was relatively low (for details see Olkowski et al. [4]). We are thus confident that any instance of a positive DENV-3 or DENV-4 result was an indication of infection with that serotype. Our base algorithm would nevertheless eliminate transient positive results for these serotypes. We, therefore, relaxed this criterion in two separate algorithms so that all positive results for DENV-3 or DENV-4 were considered a seroconversion.

Clinical serotype dominance Clinic-based surveillance data from the same period indicates that single serotypes dominated during a given transmission season. Between 2003 and 2008 the vast majority of cases were caused by DENV-3 and between 2008 and 2010 by DENV-4. Our results, however, suggest that DENV-1 and DENV-2 were continuously circulating during these times. To reconcile these differences, we evaluated two different algorithms. The first considered an individual's first seroconversion between 2003 and 2008 to be a DENV-3 infection. The second did the same for DENV-4 between 2008 and 2010. Infection with other serotypes was only possible if subsequent to a DENV-3 or DENV-4 infection the individual's serohistory showed evidence of seroconversion to DENV-1 or DENV-2.

To evaluate the impact of these alternative cleaning algorithms on the final number of seroconversions to each serotype, we implemented each one in isolation and in concert with the others, resulting in 47 alternate cleaning procedures. Those that were less conservative allowed for more overall seroconversions and appeared qualitatively similar to the main results . The least restrictive cleaning method allowed for multiple seroconversions within an interval of any length and assumed that a positive test to both DENV-3 and DENV-4 was a true seroconversion independent of later tests. This method naturally resulted in more seroconversions (Fig. S33a versus Fig. 1) and while the resulting *FoI* estimates were correspondingly larger, we obtained the same qualitative results; i.e. synchronization among serotypes.

The less restrictive cleaning methods did not address the discordance between our results and patterns observed in Iquitos clinics. Restrictions of the third type (attempting to alter the data to match clinical incidence patterns) removed a considerable number of infections to DENV-1 and DENV-2. Figure S33b plots the number of seroconversions by month when an individual's first seroconversion from 2003 to 2008 was declared DENV-3 unless the individual already appeared to have seroconverted to DENV-3. That analysis resulted in fewer DENV-1 and DENV-2 seroconversions during that interval (Fig. S33b versus Fig. 1), but importantly there were still some

seroconversions those viruses. Even under the artificially restrictive third type of cleaning regimes we could not force the cohort data to match the transmission patterns of the cohort data. This result highlights the need to consider differences between clinically apparent and inapparent transmission dynamics. We saw the same outcome for DENV-4 (Fig. S33c); there were fewer seroconversions to the other serotypes than without this restriction, but no serotype completely disappeared. Unlike estimates generated by the least restrictive cleaning method, these two cleaning methods markedly lowered our *FoI* estimates (Fig. S33b, S33d).

SECTION S3   Mathematical derivations

FoI

The simplest definition of the *FoI*, denoted λ, comes from the so called 'catalytic' model [10], where the change in the proportion of the initially susceptible class (denoted s) is governed by the following differential equation:

$$\frac{ds(t)}{dt} = -\lambda s(t) \tag{S1}$$

The solution to this differential equation is:

$$s(t) = s(t_0) \exp\{-\lambda t\} \tag{S2}$$

Perhaps the easiest way to see the effect of the *FoI* is by considering the ratio of the proportion of the study population that is still susceptible in two consecutive days. Using Eq. (S2) we have

$$\lambda = -\ln\left(\frac{s(t+1)}{s(t)}\right) \tag{S3}$$

If we consider the proportion of the initial susceptible class that have already been infected by day t (denoted $F(t) = 1 - s(t)$), we have

$$\frac{dF(t)}{dt} = \lambda(1 - F(t)) \tag{S4}$$

whose solution is

$$F(t) = 1 - \exp\{-\lambda t\} \tag{S5}$$

In the above, the *FoI* is constant. If we allow the *FoI* to vary in time (denoted $\lambda(t)$, using day as the unit of time) we can rewrite Eq. (S3) as

$$\lambda(t) \approx -\ln\left(\frac{s(t+1)}{s(t)}\right) \tag{S6}$$

Eq. (S4) as

$$\frac{dF(t)}{dt} = \lambda(t)(1 - F(t)) \tag{S7}$$

and Eq. (S5) as

$$F(t) = 1 - \exp\left\{-\int_0^t \lambda(u)du\right\} \tag{S8}$$

From Eq. (S7), writing $\frac{dF(t)}{dt} = f(t)$ and defining κ as

$$\kappa = \int_{-\infty}^{t_0} f(u)du \tag{S9}$$

where t_0 is a constant, we have

$$s(t) = 1 - F(t) = 1 - \int_{-\infty}^t f(u)du = 1 - \left(\kappa + \int_{t_0}^t f(u)du\right) \tag{S10}$$

And finally solving Eq. (S7) for $\lambda(t)$, and substituting appropriately for both $\frac{dF(t)}{dt}$ and $1 - F(t)$, we have arrived at Eq. (2).

$R(t)$ and $\mathcal{R}_0$

The *FoI* quantifies the rate at which individuals leave the susceptible pool. Our *FoI* estimates are based on, and apply to, our sample population. Under the assumption that at any moment in time the *FoI* for our sample population is the same as the population in general, the *FoI* is a function of time, but not age. To use our *FoI* estimates to calculate, for all of Iquitos, $R(t)$ and $\mathcal{R}_0$ we need to account for the difference between our sample population and the whole population. Given the presence of births, at any moment in time the age structure of the whole population will be different than that of the sample population and because individuals of different ages will have lived under the risk of dengue infection for different amounts of time, the fraction susceptible in each age class will likewise be different (even under the assumption that age has no impact on the *FoI*). If we calculate, for each age group, the fraction susceptible at any point in time and then weight those fractions by the overall fraction of the population within that age group, we can estimate the fraction of the entire population susceptible (denoted $s_P(t)$). Following [11], if we define $s_P(a, t)$ as the fraction of the individuals within the whole population that are age a and susceptible at time t, we have:

$$\frac{\partial s_P}{\partial a} + \frac{\partial s_P}{\partial t} = -\lambda(t) s_P(a, t), \tag{S11}$$

whose solution is:

$$s_P(a, t) = \exp\left\{-\int_0^a \lambda(t - a') da'\right\}. \tag{S12}$$

We assume a stable age-structure within Iquitos from 1999 through 2010 and let $p(a)$ and $P(a)$ denote the pdf and survival function of the age distribution respectively ($P(a)$ is the fraction of the population whose age is at least a, also known as the complementary cumulative distribution function). Using $s_P(a, t)$ and $p(a)$, we calculate $s_P(t)$ as:

$$s_P(t) = \int_0^\infty s_P(a', t) p(a') da', \tag{S13}$$

Note that we do not have to include a term for death within Eq. S11 or Eq. S13 because deaths are implicitly accounted for within $p(a)$.

As noted in the discussion of κ, $\lambda(t)$ can not be estimated directly from the data before t_0 (the beginning of the study period). For DENV-3 and DENV-4, because they invaded Iquitos after t_0, we can set their *FoIs* to 0 before t_0. For DENV-1 and DENV-2, we use both their respective estimates of κ and date of invasion to estimate λ before t_0. For the purpose of calculating $s_P(a, t)$ for $t > t_0$, we assume that $\lambda(t)$ is constant before t_0. For a given serotype, letting t_I represent the date of invasion, and setting $\lambda(t) = \lambda_0$ for $t_I < t < t_0$ (and 0 for $t < t_I$), we can rewrite Eq. (S12) evaluated at $t = t_0$ as:

$$s_P(a, t_0) = \begin{cases} \exp\left\{-\int_{a-(t_0-t_I)}^a \lambda_0 da'\right\}, & a \geq t_0 - t_I; \\[2em] \exp\left\{-\int_0^a \lambda_0 da'\right\}, & a < t_0 - t_I, \end{cases} \tag{S14}$$

$$= \begin{cases} \exp\left\{-(t_0 - t_I)\lambda_0\right\}, & a \geq t_0 - t_I; \\[1em] \exp\left\{-a\lambda_0\right\}, & a < t_0 - t_I. \end{cases} \tag{S15}$$

Recalling that $s(t_0)$ denotes the fraction of the sample population that is susceptible at time t_0 (i.e., those born before 1995 and thus whose age at time t_0 is at least 5) we have:

5

$$\kappa = s(t_0) \tag{S16}$$

$$= \frac{1}{P(5)} \int_5^\infty s_P(a', t_0) p(a') da' \tag{S17}$$

$$= \frac{1}{P(5)} \left(\int_5^{t_0-t_I} \exp\{-a'\lambda_0\} p(a') da' + \exp\left\{-(t_0 - t_I)\lambda_0\right\} P(t_0 - t_I) \right) \tag{S18}$$

Using the 2007 census in Iquitos [12], we estimate $p(a)$ by year. Letting $\hat{p}(a)$ denote the fraction of the population that is between a and $a + 1$ years of age and $\hat{P}(a)$ denote the fraction of the population that is at least a years old, we translate the integrals in Eq. (S18) into summations:

$$\kappa \approx \frac{1}{\hat{P}(5)} \left(\sum_{a'=5}^{\lfloor t_0-t_I-1 \rfloor} \exp\{a'\lambda_0\} \hat{p}(a') + \exp\{-(t_0 - t_I)\lambda_0\} \hat{P}\left(\lfloor t_0 - t_I \rfloor\right) \right) \tag{S19}$$

where $\lfloor \cdot \rfloor$ is the floor function. Finally, using $\hat{p}$ and the appropriate estimates of t_I and κ, we numerically solve Eq. (S19) for λ_0 for DENV-1 and DENV-2.

As discussed in the main text, the effective reproductive number, $R(t)$, is the number of secondary infections caused by an individual who is infected at time t. If we denote the force of infection on day d as $g(d)$, then we have:

$$g(d) = \int_d^{d+1} \lambda(t) dt, \tag{S20}$$

and for Iquitos (using the fact that the census indicates no individuals within Iquitos are older than 98 years of age), we have

$$s_P(t) \approx \sum_{a'=0}^{98} s_P(a', t) \hat{p}(a'). \tag{S21}$$

If N is the size of the whole population, then $N \cdot s_P(d) \cdot g(d)$ approximates the number of new infections that occured on day d. Letting $\varepsilon(d) = s_P(d) \cdot g(d)$ (i.e., the fraction of the whole population that is infected on day d) and using $w(\Delta d)$ to denote the probability that the time between two successive infections, also called the serial interval, is equal to Δd days, we can write the number of infections caused by those who became infected at time d, denoted $\pi(d)$ as

$$\pi(d) = \sum_{i=d}^\infty w(i - d) \cdot N \cdot \varepsilon(i) \tag{S22}$$

Then we can write the effective reproductive number for day d, $R(d)$, as

$$R(d) \approx \frac{\pi(d)}{N \cdot \varepsilon(d)} = \frac{\sum_{i=d}^\infty w(i - d) \cdot \varepsilon(i)}{\varepsilon(d)} \tag{S23}$$

For DENV, $w(\Delta d)$ is unknown, however recent work estimates that the most likely serial interval is 15 to 17 days [13]. As such, we approximate $w(\Delta d)$ as follows:

$$w(\Delta d) \approx \hat{w}(\Delta d) = \begin{cases} \frac{1}{3}, & \Delta d = 15, 16 \text{ or } 17; \\ 0, & \text{otherwise.} \end{cases} \tag{S24}$$

This approximation ascribes one third of infections 15, 16 and 17 days after d to the infections that occurred on day d. Then, by approximating $w(\Delta d)$ in Eq. (S23) by $\hat{w}(\Delta d)$ from Eq. (S24), denoting the approximate value of $R(d)$ as $\hat{R}(d)$, and using the substitution given in Eq. (S20) we derive Eq. (5). An alternative derivation of $R(t)$

follows from [14]. There, using a continuous version of $w(\Delta t)$ to represent the distribution of the serial interval, the substitution of a continuous $\hat{w}(\Delta t)$ analogous to Eq. (S13) results in a similar representation for $R(t)$. Finally, we relate $\hat{R}(d)$ to $\mathcal{R}_0$ through the following equation (where again $s_P(d)$ is the fraction of the whole population susceptible on day d)

$$\hat{R}(d) = s_P(d)\mathcal{R}_0 \tag{S25}$$

Solving Eq. (S25) for $\mathcal{R}_0$, we arrive at Eq. (6).

SECTION S4   MCMC description, convergence checks, model selection and identifiability

Adaptive, Metropolis-within-Gibbs MCMC algorithm

Due to the number of parameters we estimated, as well as the difficulty of computing the required conditional distributions to use a Gibbs sample, we used a Metropolis-within-Gibbs algorithm [15]. For each step in the chain this algorithm cycles through each individual parameter, and proposes updating that parameter's value using the Metropolis algorithm. Specifically, a single draw from a one-dimensional Normal distribution (to be explicitly defined below) is added to the parameter, the likelihood is computed for the new set of parameters (say L^*, with the likelihood before the perturbation denoted L^{old}), and the new proposed parameter is accepted with probability, p, where p is defined as:

$$p = \begin{cases} \frac{L^*}{L^{\text{old}}}, & \text{if } L^* < L^{\text{old}}; \\ \\ 1, & \text{otherwise.} \end{cases} \tag{S26}$$

Every parameter (order randomly chosen), is potentially perturbed using the same rule (but not necessarily the same Normally distributed noise), and at the completion of the proposals, a single step of the MCMC chain has been completed.

To decrease the time the chains take to converge to computationally reasonable times, we also incorporate an adaptive MCMC algorithm [16] (the adaptive algorithm is discussed within the reference). The adaptive aspect of the algorithm concerns the standard deviation of the distribution of the proposal noise. Optimally [17, 18], one would draw the noise for the proposals from the distribution

$$N\left(0, \frac{(2.38)^2 \Sigma}{d}\right) \tag{S27}$$

where Σ denotes the covariance structure of the target distribution and d is the number of parameters. Because Σ is unknown, the adaptive MCMC algorithm approximates it from the empirical estimate based on a number of draws.

From an initial estimate of κ and the 72 B-spline coefficients we generated an initial approximation of f. Note that to assess the robustness of our results when starting each parameter of each chain at the same initial value (set to 0.001), we repeated the analysis so that every parameter of every chain was chosen from a uniform [0,0,001] distribution (Fig. S12). There was no appreciable difference between the posterior distributions of the two approaches. Using this initial f, we computed an initial likelihood. We initially started our MCMC chain using the d-dimensional identity matrix, I_d, as the null estimate, Σ_0, of Σ. Then we selected one of the 73 parameters at random, perturbed it by adding Normally distributed noise, and chose to accept the new value of the parameter using the Metropolis-Hastings acceptance rule. Once every parameter was perturbed once (and either the new value was adopted, or the parameter was reverted to its original value), the chain completed a single step. After an initial run of 5000 steps with independent Gaussian noise, we updated this estimate with the empirical covariance of those 5000 runs (plus, as recommended in [16], a small amount of 'nonrandom normal' as a "safety measure" (see Eq. 2.1 in [16])). We ran the Metropolis-within-Gibbs algorithm with this proposal distribution for 5,000 more steps

and then repeated the adaptive step using the last 5,000 steps, and continued. We did this twice more, for a total of 4 adaptive steps, and a total of 20,000 steps. Finally, we then ran the Metropolis-within-Gibbs algorithm with the 5th estimate of Σ for 80,000 more steps. In this final step, we used only the diagonal elements from Σ to re-achieve Markovian chain-steps.

Convergence was assessed through several metrics. We retained the final 15,000 steps for our analysis. We then recombined the steps from 10 different converged chains to arrive at 150,000 draws from the posterior distribution. From these final 150,000 steps, we randomly sampled 1,000 steps from this subset of the chain to remove autocorrelation, and then using Eq. (3) and the mean value of each parameter among those 1,000 steps, we obtained an estimate of f through time. Because the parameters were not independent of each other, to create credible regions for f (and later the *FoI* and $\mathcal{R}_0$), we used the 1,000 sampled steps of the chain to create 1,000 estimates of f. This forms an empirical estimate of the posterior distribution of f. We then, for each day, selected the middle 90% of the estimates to form our credible interval at that point.

Assessing convergence of MCMC chains

We generated 10 chains independently for each set of parameters with the intent to combine them and form one final sequence of parameter sets. In an attempt to ensure convergence of these chains (and due to the relatively quick time a single step in the chain takes to compute), we ran them each 80,000 steps after the final adaptive step. 80,000 was not chosen to minimize the number of steps required before convergence was declared, but rather as an attempt at overkill because convergence can never be truly 'confirmed.' Convergence checks at lower chain lengths appeared to converge, but there was no reason not to compute excessively long chains to increase our confidence in convergence. The convergence checks we used were: (1) visual inspection of the values of multiple parameters as they changed through the chain (trace plot), (2) investigation of proposal noise by calculating parameter-specific empirical estimates of the acceptance probabilities, (3) calculation of Gelman and Rubin's potential scale reduction factor [19], and (4) calculation of Gelman and Brooks' multivariate scale reduction factor [20]. Because we are combining multiple independent chains, these diagnostics seem most appropriate.

For both the potential scale reduction factor of each parameter, as well as the multivariate scale reduction factor, values considerably larger than one indicate a lack of convergence (of either the individual parameter or the entire parameter set as a whole respectively). For each serotype's estimation of f and κ, there were 73 parameters within the model. Below, in Fig. S13, we plotted the potential scale reduction factor (PSRF) for each parameter, by serotype, as the chain progresses. As noted in [20], when a PSRF was below 1.1 we may have believed the parameter has converged, but this low value may be transient and monitoring this value over time increases our confidence that the parameter has converged. For DENV-1, DENV-2, and DENV-3 (Fig. S13a, b, and c respectively), we saw that the PSRF was consistently below 1.1 for all parameters by the 10,000th step in the chain (and were all below 1.02). Interestingly, the PSRF for DENV-3 takes the longest to converge for the parameters that govern f around where the serotype was introduced into Iquitos. For DENV-4 (Fig. S13d), we saw that there were considerably more steps required before we would believe convergence occurred. The PSRF for parameters 60-70 were not consistently below 1.1 until the 30,000th iteration, and for a few parameters it was above 1.02 when the chain ended. The multivariate scale reduction factors were 1.01 for DENV-1, 1.01 for DENV-2, 1.02 for DENV-3 and 1.03 for DENV-4. Again, DENV-4 had the largest factor, but all were well below 1.1 indicating a high level of confidence that the chains have converged.

We also assessed convergence using trace plots and by evaluating the acceptance probabilities. Because there were 73 parameters, we present only a sampling of six of the trace plots (Fig. S14) and acceptance probabilities (Fig. S15). We combined the last 15,000 steps (indicated by the vertical dashed lines in Fig. S14) of 10 chains, but for clarity we have only displayed the values for the first 3 chains. For DENV-4, because its invasion was not until 2008, we selected a different subset of parameters to plot, starting at the 35th parameter. As indicated by

the PSRF analysis, the mixture of some later parameters for DENV-4 (e.g., the 63rd parameter in Fig. S14d) was sub-optimal. The estimated acceptance probabilities were running averages of the ratio of accepted proposals divided by the total number of values proposed. Our adaptive step ends after 20,000 steps and as such, we did not achieve perfect acceptance probabilities for every parameter of every chain for every serotype. Parameters that have relatively poor mixture (e.g., the 63rd parameter of DENV-4) also appeared to have lower than desirable acceptance probabilities (see, e.g., the 63rd parameter of DENV-4, Fig. S15d). As individual chains, they did not all appear appropriate for inference. When considering the set of 10 chains simultaneously, however, the resulting collection of parameters appears adequately converged.

Model selection

The deviation information criterion (DIC) [21] is a measure of a model's out-of-sample predictive power, and as such can be used to select the optimal model. We define the deviance, with respect to the data y and the parameter set θ, (denoted as $D(y, \theta)$) as:

$$D(y, \theta) = -2 \log p(y|\theta) \tag{S28}$$

where $p(y|\theta)$ is the likelihood function. Further, over the set of parameter sets that constitute the posterior distribution obtained through MCMC, we evaluate the deviance for each parameter set and denote their average as $\overline{D}$. Finally, letting $\hat{\theta}$ to be the mean parameter set of the posterior simulations, the DIC is defined as:

$$DIC = 2\overline{D} - D(y, \hat{\theta}) \tag{S29}$$

Because we analyzed 12 years of data, we chose to use our model selection procedure to select the number of splines per year, not in total. We did not alter the locations of the knots that define the centers of the splines from their default positions and as such these two approaches (selecting the number of splines per year versus total number of splines) were analogous. For the purposes of comparing fitted models across serotypes, we used the same number of splines per year for each serotype. As such, we identified the optimal number of splines for each serotype's model and selected the maximum of those for the final set of models. It is important to note that because DENV-4 was not tested for until 2006, the total number of splines used was the selected number of splines per year multiplied by 5 years (as opposed to 12 years for DENV-1, DENV-2, and DENV-3).

From Table S2 3, 4 and 6 splines per year all produced comparable DIC values for DENV-1 and DENV-2, with the model with 4 splines being most preferred. For DENV-3, 6 splines per year was significantly better than any other model, and for DENV-4, 3 splines a year was optimal. As such, we selected 6 splines per year for each model, and acknowledge (especially given the relatively flat fit for DENV-4) that fewer splines would have been sufficient for DENV-4. Because DIC indicated that 4 splines per year was adequate for 3 of the 4 serotypes, we conducted all analyses with 4 splines (Section S6).

Identifiability issues

There were particular continuity issues that occurred at the boundaries of the region over which the B-splines were defined. In the middle of this region, splines were placed at evenly spaced knots. Here, for comparability across serotypes, we did not choose to optimize the location of the interior spline knots by placing more of them where more interval infections were observed. Instead, for all serotypes, knots were placed unevenly at both endpoints (specifically they were spaced closer than usual). The relative effect of the last spline on the function f was minimal. Because there was not much data from the end of the study, it was not surprising that this parameter would be poorly identified. For the purposes of the analysis, we accounted for this difficulty by truncating our estimates accordingly. Graphically, we removed the final 60 days, but within the likelihood we retained this parameter.

The posterior distribution of the final spline parameter was quite wide (Fig. S16). Across serotypes, the median of the posterior for this parameter was considerably larger than for any other parameter. When looking at the actual likelihoods across values of this parameter, the likelihood was maximized when this parameter was essentially zero. Note that the solid lines in Figure S16 were computed holding all other parameters constant at their median posterior value and thus this was not the profile likelihood. The profile likelihood would be flatter, but the maximum would remain in the same place.

In addition to the endpoint issue, another identifiability issue arose for the first spline parameter. The likelihood integrates f and then adds the integral to κ. Thus the value of the first spline parameter was naturally additively confounded with the value of κ; if one was lower, the other could be increased to create essentially identical likelihoods. To illustrate the relationship between the first spline parameter and κ, we plotted the two dimensional posterior distribution of these parameters in Figure S17. For DENV-3 and DENV-4 (Fig. S17c,d respectively), both of the parameters should be exactly 0 and the fact that they weren't 0 was an artifact of the fitting procedure. Although the values were quite small, we still saw the appearance of a linear relationship between the two parameters in that when one was infinitesimally larger, the other was infinitesimally smaller. The relationship was clearer for DENV-1 and DENV-2 (Fig. S17a, b respectively). There was a clear linear relationship between the two parameters. This variation in κ was proportionally small, but the effect of this relationship resulted in considerable increases in the first spline parameter. To avoid over-interpretation of this effect (which, when confronted with our data, was clearly an artifact of our fitting procedure), we truncated the first 100 days' output in the beginning of f for our analysis.

In Figure S19, we plotted, by serotype, our estimates of f. On each panel we indicated the periods of time that we truncated due to the indentifiability issues. Comparing this plot to Figure 1, it is clear that there were no data to support the excessively high values at the beginning and end of each estimate. The magnitude of f was small, and thus the identifiability issue greatly skewed our estimates in the indicated regions.

SECTION S5   Additional results

Due to a limited amount of space in the main text and the multitude of almost equally-critical products of our analysis, we present here additional results.

κ, $s(t)$ and $s_P(t)$

The other outputs of our model are, by serotype, the fraction of the study population that was infected before 1999, κ. In Figure S18 we plotted the posterior distributions of κ, again noting non-zero values for DENV-3 and DENV-4 (Fig. S18 c and d respectively) were artifacts of the fitting method. Using Eq. S10, we combined f and κ to compute, by serotype, the fraction of the study population susceptible at time t, $s(t)$ (Fig. S20). There was a relatively steady decline in susceptible individuals in the two previously circulating serotypes (DENV-1 (orange) and DENV-2 (green)). DENV-3 (blue) and DENV-4 (purple) both experienced sharp declines consistent with the initial outbreaks of a pathogen in a wholly susceptible study population. Relative to the other serotypes, DENV-4 appeared to have been somewhat consistently depleting its pool of susceptible individuals since introduction. Finally, using λ, κ and $\hat{p}(a)$ we estimate $s_P(t)$. The patterns for DENV-3 and DENV-4 are relatively similar (Fig. S21). Conversely, once the differences between the sample population and all of Iquitos were accounted for (namely the addition of new suceptibles through births), the fraction of the population that was susceptible to DENV-1 and DENV-2 are mostly stable. Over the 12 years, the median fraction of the population that is susceptible to DENV-1 and DENV-2 varies by 8% and 6% respectively (36.6%-44.6% for DENV-1 and 40.0%-46.0% for DENV-2).

Robustness of estimates to the 2005-2006 gap

For our primary algorithm, we did not adjust for the gap in cohorts from late-2005 to mid-2006. This gap afforded us an opportunity to investigate the robustness of our estimates by providing a natural splitting point of the data. For this analysis, we cut the data in half (half 1: 1999-2005, half 2: 2006-2010) and performed our analyses on each half separately. The resulting estimated FoIs are plotted in Figure S24. There was considerable agreement between these estimates and our original estimates. There were, however, some noticeable differences. On both sides of the gap, the estimated credible intervals and posterior medians were considerably larger. Because the original estimates were all dependent on each other, the original estimates in mid-2005 were informed by all the data from 2006-2010, and likewise the original estimates in 2006 were informed by the data from 1999-2005. Thus, it was not too surprising that the credible intervals and estimates were larger.

$R(t)$, $\mathcal{R}_0(t)$, $\mathcal{R}_0(t)$'s sensitivity to the serial interval and $\mathcal{R}_0(t)$'s relationship to *FoI*

$\mathcal{R}_0$ is defined as the number of secondary infections generated by a single infectious individual entering an entirely susceptible population. In practice, this is difficult to observe. Conversely, $R(t)$, the effective reproductive number, merely counts the number of secondary infections at any point in time, independent of the number of susceptibles. In Figure S26 we plotted the estimated daily effective reproductive number for each serotype. The highest computed values of $R(t)$ were for DENV-3 in 2002 where it surpassed 3.

As mentioned in the main text, the presence of s_P in the denominator of $\mathcal{R}_0(t)$ allowed the tail of the posterior distribution of $\mathcal{R}_0(t)$ to be quite fat. In the main text we compensated for this by displaying the 50% credible interval from the posterior distribution of $\mathcal{R}_0(t)$ (Fig. 5), but in Figure S27, to be consistent with the credible level displayed in other figures, we display the fitted values of $\mathcal{R}_0(t)$, as well as the 90% credible interval from the posterior distribution. Note that the scale of $\mathcal{R}_0(t)$ had to be considerably increased to allow for the rare, but extremely high values at the upper end of the credible interval. For example, in 2007 and 2010 the upper limit of the 90% credible interval for the $\mathcal{R}_0(t)$ of DENV-3 exceeded 20. When we smoothed the data by computing a weekly $\mathcal{R}_0(t)$ (by taking the number of estimated cases in one week and then the number of cases 15 to 24 days into the future (appropriately scaling)), this extreme behavior disappears, and we preserved the general pattern. Because the time scale of f was daily, we chose to remain consistent across all products that use f in the main text and figures.

To compute the yearly average estimates for $\mathcal{R}_0$ (Fig. S32), within each year we took a weighted average of the daily $\mathcal{R}_0$ values. We weighted these daily values by the relative number of infectious individuals for each of those days. If we denote $\overline{R_0(Y)}$ the yearly average estimate for year Y, we have

$$\overline{R_0(Y)} = \frac{1}{\sum_{t \in Y} f(t)} \sum_{t \in Y} f(t) \cdot R_0(t) \tag{S30}$$

The serial interval between successive infections we used was 15 to 17 days [13]. This resulted in averaging the number of infections over those three days when computing the effective reproductive number. To assess the sensitivity of our results to the particular serial interval we used, we repeated the computation using shorter (just 15 days) and longer serial intervals (15 to 19 days). In both cases, we assumed the distribution of the serial interval was uniform across the interval to compute $\mathcal{R}_0$. In Fig. S28 we plotted estimates for shorter (Fig. S28a) and longer (Fig. S28b) serial intervals. The results were visually comparable, and numerical investigation showed that $\mathcal{R}_0$ estimates were at most 9% lower and 6% higher than those based on the 15-17 day serial interval.

A complete transmission cycle of DENV requires passage through two different latent periods. Starting with an infectious human host, a susceptible mosquito takes a virus-infected blood meal, and the pathogen enters an 'extrinsic incubation period' (EIP) in the mosquito. There are several different published estimates of the duration of the EIP ranging from 7-13 days [22, 23, 24, 25]. Following the EIP, the mosquito is infectious for the remainder of its life. After the infectious mosquito takes a blood meal from a susceptible human host and successfully infects this host, the virus enters an 'intrinsic incubation period' (IIP). There are several different published estimates of

the IIP, ranging from 4-7 days [26, 22, 23, 27, 28]. Using the shortest combination of EIP and IIP estimates, we estimated the shortest possible duration of the serial interval to be 11 days, which is consistent with the empirical data Siler et al reported [26]. We conducted two additional sensitivity analyses of $\mathcal{R}_0$, using first 11-13 days (Fig. S29a) and then combining that minimum (11) with the maximum presented by Aldsadt et al [13] (17), resulting in a relatively wide range for the serial intervals ranging from 11-17 days (Fig. S29b). The results were also visually comparable with slightly lower global and local maxima for estimates based on serial intervals with shorter lower bounds. Numerical investigation showed that $\mathcal{R}_0$ estimates were at most 18% lower and 16% higher than those based on the 15-17 day serial interval.

Also as mentioned in the text, there appeared to be a lag between $\mathcal{R}_0(t)$ and the *FoI* (Fig. S31). In Figure S30, we plotted the correlation (both Pearson (Fig. S30a) and Spearman (Fig. S30b)) between $\mathcal{R}_0(t)$ and variously lagged values of the *FoI*. For both, we used the median estimated daily lag value. In both panels it is clear that the highest correlation was between $\mathcal{R}_0(t)$ and the *FoI* of DENV-1 (orange), but there were relatively high maximum correlations for each serotype. The timing of the maximized correlation for each serotype was between 68 and 75 days for the Pearson correlation and between 64 and 72 days for the Spearman correlation.

Due to the number of B-splines per year used, we also investigated the level of uncertainty of both weekly and monthly estimates of the *FoI* and $\mathcal{R}_0$. In both cases (Fig. S25), the credible intervals remained relatively constant. For the *FoI* the values increased as expected but maintained the patterns of the daily estimates. For $\mathcal{R}_0$, the maximum values decreased slightly for weekly values and monthly values. (Fig. S25c, S25d respectively). For the weekly and monthly $\mathcal{R}_0$ estimates, like the yearly estimates, we took weighted averages of the daily estimates within the corresponding ranges.

Forward simulation using $\lambda(t)$

Using our estimates of λ, κ and $\hat{p}(a)$, we can forward simulate age-structured serotype specific DENV infection dynamics within Iquitos (assuming no interactions between serotypes). Let s_a, e_a, i_a and r_A denote the fraction of individuals between a and $a+1$ years old that is susceptible, exposed, infectious and immune respectively. For each DENV serotype, we can model the transmission dynamics forward using a standard SEIR model with age dependence. In particular, using a system of difference equations, our susceptibility estimates as initial conditions, and our daily estimates of $\lambda(d)$ to drive transmission, we have:

$$s_a(d+1) = \frac{s_{a-1}(d)}{365} + \frac{364}{365}s_a(d) - \lambda(d)s_a(d) \tag{S31}$$

$$e_a(d+1) = \frac{e_{a-1}(d)}{365} + \frac{364}{365}e_a(d) + \lambda(d)s_a(d) - \alpha e_a(d) \tag{S32}$$

$$i_a(d+1) = \frac{i_{a-1}(d)}{365} + \frac{364}{365}i_a(d) + \alpha e_a(d) - \gamma i_a(d) \tag{S33}$$

$$r_a(d+1) = \frac{r_{a-1}(d)}{365} + \frac{364}{365}r_a(d) + \gamma i_a(d) \tag{S34}$$

where

$$S(d) = \sum_{a'=0}^{98} s_{a'}(d)\hat{p}(a'), \tag{S35}$$

$$I(d) = \sum_{a'=0}^{98} i_{a'}(d)\hat{p}(a'), \tag{S36}$$

$$s_{-1}(d) = 1 \quad \text{for all } d, \tag{S37}$$

$$e_{-1}(d) = i_{-1}(d) = r_{-1}(d) = 0 \quad \text{for all } d, \tag{S38}$$

$1/\gamma$ is the average infectious period, and $1/\alpha$ is the average latent period. To match our serial interval estimates, we approximate $1/\gamma$ at 15 days. Following historical estimates of the average infectious period [29], we set $1/\alpha$ to 4 days.

Using the median output from our six B-Spline per year model, we deterministically simulate forward the age-structured SEIR model. In Figure S22 the estimated total number of infectious individuals within Iquitos by age and day is plotted (calculated by multiplying $I(d)$ by the population size of Iquitos). Similar to Figure 3, the high level of immunity for DENV-1 and DENV-2 limits the total number of infectious individuals on any day. As such, the majority of infections that occur are with young children. The estimated age-distribution of DENV infections for October 1, 2008 (indicated by the red line in Fig. S22) is plotted separately in Figure S23. This date is near the estimated invasion of DENV-4 and as such the age-distribution of infectious individuals is almost exactly the same as the age-distribution of individuals within Iquitos (Fig. S23d indicated by the dark grey histogram). Younger children are over-represented in DENV-1 and DENV-2 infections while high levels of immunity cause adults to be under-represented (Fig. S23a,b). DENV-3 infections display slightly different age-distribution patterns than DENV-1 and DENV-2 (Fig. S23c). Although adults are still under-represented, there are proportionally more individuals over the age of 10, indicative of a serotype that has not stabilized within the population.

SECTION S6   Four B-Splines per year

To assess the robustness of our results regarding the decision to use 6 B-splines per year for our estimation algorithm, we recreated all of the primary analyses using 4 B-splines per year. This resulted in a model with 49 parameters per serotype. We used an identical MCMC approach, initializing every chain individually by selecting random initial conditions.

The *FoI*

Analogously to Fig. 3 and Fig. 4 for the 4 spline-per-year model we present Fig. S2 and Fig. S3. The daily results for DENV-1 and DENV-2 were almost identical to those computed using the 6 spline-per-year model. Results were different for DENV-3. Because the DIC for DENV-3 was over 20 points lower for the 6 spline-per-year model, it was not surprising that there was some level of discordance between the two models' output here. At the yearly scale, all results were qualitatively identical. As with the 6 spline-per-year model, there was a high level of correlation between the yearly DENV-1, DENV-2 and DENV-3 estimates for the 4 spline-per-year model. Here the Spearman correlations were $\rho_{12} = 0.78$; DENV-1/DENV-3: $\rho_{13} = 0.73$ and DENV-2/DENV-3: $\rho_{23} = 0.79$, which were very similar to those described in the main text for the 6 spline-per-year model.

$R(t)$ and R_0

Analogous to Fig. S26 and Fig. 5, for the 4 spline-per-year model we present Fig. S6 and Fig. S4. In Fig. S6, it is clear that the fewer splines resulted in slightly smoother estimated daily effective reproductive numbers. The periods when the *FoI* was different for DENV-3 between the 6 spline and the 4 spline model also resulted in differences in R(t). For R_0, the maximum estimated value was for DENV-1 in 2009 (4.27). Also analogous to Fig. S32, we computed the yearly estimate of the average value of R_0 (Fig. S5). As with the daily values these were slightly lower, but overall in agreement with, the 6 spline-per-year model. The maximums of the yearly estimates were significantly lower than the maximums of the daily estimates with the maximum yearly R_0 (DENV-1 in 2010), never exceeding 3.

For R_0 we computed serial interval sensitivity analyses for the 4 spline-per-year model as outlined in SI 4. Conclusions from 6 spline-per-year and 4 spline-per-year models were essentially identical when the serial interval was set at 15 days and between 15 and 19 days (Fig. S9a, b respectively). As with the 6 spline-per-year output,

decreasing the lower bound of the serial interval to 11 days resulted in slightly lower estimates than in the primary analysis (Fig. S10).

κ, and $s(t)$

As with the 6 spline-per-year model, the estimate of κ for DENV-3 and DENV-4 for the 4 spline-per-year model were essentially 0 (Fig. S8c,d). For DENV-1 the estimated fraction of the sample population susceptible in 1999 was 56.3% (Fig. S8a) and 53.5% of the population was estimated susceptible to DENV-2 in 1999 (Fig. S8b). These values were slightly higher than those for the 6 spline-per-year model but both were well within the credible intervals reported in the main text. The trajectory of the fraction of the population susceptible to each serotype over time was similar to that of the 6 spline-per-year model (Fig. S7).

References and Notes

[1] Halstead, S. B, Rojanasuphot, S, & Sangkawibha, N. (1983) Original antigenic sin in dengue. *The American Journal of Tropical Medicine and Hygiene* **32**, 154–156.

[2] Kuno, G, Gubler, D, & Oliver, A. (1993) Use of 'original antigenic sin' theory to determine the serotypes of previous dengue infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **87**, 103–105.

[3] Morrison, A, Minnick, S, Rocha, C, Forshey, B, Stoddard, S, Getis, A, Focks, D, Russell, K, Olson, J, Blair, P, Watts, D, Sihuincha, M, Scott, T, & Kochel, T. (2010) Epidemiology of dengue virus in iquitos, peru 1999 to 2005: interepidemic and epidemic patterns of transmission. *PLoS Neglected Tropical Diseases* **4**, e670. 10.1371/journal.pntd.0000670.

[4] Olkowski, S, Forshey, B. M, Morrison, A. C, Rocha, C, Vilcarromero, S, Halsey, E. S, Kochel, T. J, Scott, T. W, & Stoddard, S. T. (2013) Reduced risk of disease during postsecondary dengue virus infections. *Journal of Infectious Diseases* pp. 1026–1033.

[5] Morens, D, Halstead, S, Repik, P, Putvatana, R, & Raybourne, N. (1985) Simplified plaque reduction neutralization assay for dengue viruses by semimicro methods in BHK-21 cells: comparison of the BHK suspension test with standard plaque reduction neutralization. *Journal of Clinical Microbiology*.

[6] Comach, G, Blair, P. J, Sierra, G, Guzman, D, Soler, M, Quintana, M. C. d, Bracho-Labadie, M, Camacho, D, Russell, K. L, Olson, J. G, & Kochel, T. J. (2009) Dengue Virus Infections in a Cohort of Schoolchildren from Maracay, Venezuela: A 2-Year Prospective Study. *Vector-Borne and Zoonotic Diseases* **9**. 10.1089/vbz.2007.0213.

[7] Kochel, T. J, Watts, D. M, Halstead, S. B, Hayes, C. G, Espinoza, A, Felices, V, Caceda, R, Bautista, C. T, Montoya, Y, Douglas, S, et al. (2002) Effect of dengue-1 antibodies on American dengue-2 viral infection and dengue haemorrhagic fever. *The Lancet* **360**, 310–312.

[8] Liebman, K, Stoddard, S, Morrison, A, Rocha, C, Minnick, S, Sihuincha, M, Russell, K, Olson, J, Blair, P, Watts, D, Kochel, T, & Scott, T. (2012) Spatial dimensions of dengue virus transmission across interepidemic and epidemic periods in Iquitos, Peru (1999-2003). *PLoS Neglected Tropical Diseases* **6**, e1472. 10.1371/journal.pntd.0001472.

[9] Thomas, S. J, Nisalak, A, Anderson, K. B, Libraty, D. H, Kalayanarooj, S, Vaughn, D. W, Putnak, R, Gibbons, R. V, Jarman, R, & Endy, T. P. (2009) Dengue plaque reduction neutralization test (PRNT) in primary and secondary dengue virus infections: how alterations in assay conditions impact performance. *The American Journal of Tropical Medicine and Hygiene* **81**, 825.

[10] Muench, H. (year?) Derivation of rates from summation data by the catalytic curve. *Journal of the American Statistical Association* **29**, 25–38.

[11] Ferguson, N, Donnelly, C, & Anderson, R. (1999) Transmission dynamics and epidemiology of dengue: insights from age-stratified sero-prevalence surveys. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* **354**, 757–768. 10.1098/rstb.1999.0428.

[12] Instituto Nacional de Estadistica e Informatica. (2007) Censos nacionales 2007 (http://http://censos.inei.gob.pe/Censos2007/redatam/).

[13] Aldstadt, J, Yoon, I.-K, Tannitisupawong, D, Jarman, R, Thomas, S, Gibbons, R, Uppapong, A, Iamsirithaworn, S, Rothman, A, Scott, T, & Endy, T. (2012) Space-time analysis of hospitalised dengue patients in rural thailand reveals important temporal intervals in the pattern of dengue virus transmission. *Tropical Medicine & International Health : TM & IH* **17**, 1076–1085. 10.1111/j.1365-3156.2012.03040.x.

[14] Wallinga, J & Teunis, P. (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology* **160**, 509–516. 10.1093/aje/kwh255.

[15] Tierney, L. (1994) Markov chains for exploring posterior distributions. *The Annals of Statistics* **22**, 1635 – 2134.

[16] Roberts, G. O & Rosenthal, J. S. (2009) Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics* **18**, 349–367. 10.1198/jcgs.2009.06134.

[17] Roberts, G, Gelman, A, & Gilks, W. (1997) Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7**, 110 – 120.

[18] Roberts, G & Rosenthal, J. (2001) Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science* **16**.

[19] Gelman, A & Rubin, D. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.

[20] Stephen, P. B & Andrew, G. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455. 10.1080/10618600.1998.10474787.

[21] Gelman, A, Carlin, J. B, Stern, H. S, & Rubin, D. B. (2004) *Bayesian Data Analysis*. (Chapman & Hall/CRC).

[22] Newton, E & Reiter, P. (1992) A model of the transmission of dengue fever with an evaluation of the impact of ultra-low volume (ULV) insecticide applications on dengue epidemics. *The American Journal of Tropical Medicine and Hygiene* **47**, 709–720.

[23] Focks, D, Daniels, E, Haile, D, & Keesling, J. (1995) A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *The American Journal of Tropical Medicine and Hygiene* **53**, 489–506.

[24] Favier, C, Schmit, D, Mller-Graf, C, Cazelles, B, Degallier, N, Mondet, B, & Dubois, M. (2005) Influence of spatial heterogeneity on an emerging infectious disease: the case of dengue epidemics. *Proceedings of the Royal Society B* **272**, 1171–1177. 10.1098/rspb.2004.3020.

[25] Otero, M, Barmak, D, Dorso, C, Solari, H, & Natiello, M. (2011) Modeling dengue outbreaks. *Mathematical Biosciences* **232**, 87–95. 10.1016/j.mbs.2011.04.006.

[26] Siler, J, Hall, M, & Kitchens, A. (1926) Dengue: Its history, epidemiology, mechanism of transmission, etiology, clinical manifestations, immunity and prevention. *Philippine Bur. Sci.*

[27] Gubler, D & Kuno, G. (1997) *Dengue and dengue hemorrhagic fever*. (Cab International Wallingford, New York).

[28] de Castro Medeiros, L, Castilho, C, Braga, C, de Souza, W, Regis, L, & Monteiro, A. (2011) Modeling the dynamic transmission of dengue fever: investigating disease persistence. *PLoS Neglected Tropical Diseases* **5**. 10.1371/journal.pntd.0000942.

[29] Gubler, D, Suharyono, W, Tan, R, Abidin, M, & Sie, A. (1981) Viraemia in patients with naturally acquired dengue infection. *Bulletin of the World Health Organization* **59**, 623–630.

Study	Time Span	Population	Samples	Study Design	Active surveillance component	IRB protocol numbers and approval*	Related publications	Serotypes tested
Entomological Correlates of Dengue Control (ECDC)	Jan. 1999 to Mar. 2005	Primary school children (78%, 5-17 y/o)	N = 3,903 with 2-12 samples at ~6 mo intervals	Participants selected from geographically stratified sample of city blocks spanning northern half of Iquitos	School-based in subset of participants	UCD-235832, NM-RCD2001.008, INS13-2002 SDSU	Morrison et al. 2004a,b, 2010, Rocha et al. 2009	DENV-1, -2, -3
Dengue Vector Control System (DVCS)	April 2004 to Oct. 2005	Ages 5 and above (77%, >18 years)	N= 1,267 with 2-3 samples at 9 mo intervals	Participants were enrolled from 24 city blocks distributed across 7 of 26 administrative zones located in the northern half of Iquitos	Community-based surveillance, 3 visits per week	UCD-200311958, NMRCD2003.0008, INS-04502003	Rocha et al. 2009	DENV-1, -2, -3
Predictors of Disease Severity (PRED)	August 2006 to June 2010	Ages 5 and above (66%, >18 years)	N=2,555 with 2-7 samples at 6 mo intervals	Cohort members were recruited from 20 city blocks selected (2 per zone) at random from 10 administrative zones in northern half of Iquitos	Community-based surveillance, 3 visits per week	UCD-216811 NM-RCD2005.0009 CAY-06017	Forshey et al. 2010 (Influenza)	DENV-1, -2, -3, -4
Measuring Risk across Activity Spaces (AS)	November 2007 to December 2010	Ages 5 and above (76%, >18 years)	N=3,500 with 2-5 samples at 6-12 mo intervals	Cohort members were recruited from contiguous blocks in two neighborhoods: Maynas (20 blocks) and Tupac Amaru (14 blocks).	Community-based surveillance, 3 visits per week	UCD-296683 NM-RCD2007.0007 Relying Agreements: TUL, EM	Stoddard et al. 2009, 2013 Vasquez-Prokopec et al. 2009 Paz-Soldan et al. 2010	DENV-1, -2, -3, -4
Insecticide Treated Curtains (ITC)	October 2009-August 2010	Ages 3 and above (71%, >20 years)	N=1,943 with 2 samples 9 mo interval	Cohort was recruited from 20 clusters (90 houses) from a contiguous area located in southern Iquitos	No surveillance component	NMRCD.2009.0007 LSTM, LSTMH Relying Agreements: UCD, TUL		DENV-1, -2, -3, -4

*Institutional Review Board (IRB) abbreviations. University of California (UCD). Naval Medical Research Center Detachment (NMRCD) or Naval Medical Research Unit No. 6 (NAMRU-6) where from 1999-2007 protocols were reviewed by the Naval Medical Research Center (NMRC) IRB located in Bethesda Maryland and assigned NMRCD numbers to protocols carried out by NMRCD which became NAMRU-6 in July 2010. In 2007, NMRCD which later became NAMRU-6 formed their own IRB which is comprised of both Peruvians and military officials and because of its registration with the Peruvian Network of Ethic's Committee provides both DOD and Peruvian approval for reviewed protocols. Studies conducted before 2007 were reviewed by the NMRC and either the Peruvian National Institute of Health (INS) or Cayetano Heredia University (CAY) located in Lima, Peru to obtain Peruvian approval. Additional collaborators, Tulane University (TUL), Emory University (EM), San Diego State University (SDSU) established relying agreements with either UCD or NMRCD/NAMRU-6 IRBs. The ITC project was also approved by the Liverpool School of Tropical Medicine (LSTM) and London School of Tropical Medicine and Hygiene (LSTMH) at the initiation of the study as per European Regulations.

Table S1: **Summary of cohort studies.** For each longitudinal cohort study, we list the time span, population of interest, total number of participants/samples, a brief description of the study design, a brief description of the active surveillance component, IRB protocol numbers, related publications, and the DENV serotypes tested for.

	Number of splines per year				
	2	3	4	6	8
DENV-1	19930.11	19908.29	19904.41	19910.37	19931.03
DENV-2	21009.92	21005.29	21003.36	21007.11	21028.16
DENV-3	23461.92	23397.27	23356.47	23335.02	23342.9
DENV-4*	11535.79	11528.3	11537.6	11551.92	11572.8

* Due to DENV-4 not being tested until 2006, the total number of spline parameters is based on 5 years.

Table S2: **Summary of DIC values.** For each serotype, we utilized deviation information criterion (DIC) to select the appropriate model. For each serotype, we list the DIC value for the model that consisted of 3, 4, 6, or 8 splines per year. For comparability across serotypes, we chose, amongst the recommended models for each serotype, the largest model which resulted in our models each having 6 splines per year.

$P(T < t_L) = \int_{-\infty}^{t_L} f(s)ds$	For left censored individuals who seroconverted before their first observation at time t_L.
$P(T_{I_1} < I < t_{I_2} = \int_{t_{I_1}}^{t_{I_2}} f(s)ds$	For interval censored individuals who seroconverted between observation times t_{I_1} and t_{I_2}.
$P(T > t_R) = 1 - \int_{-\infty}^{t_R} f(s)ds$	For right censored individuals who never seroconverted and were last observed at time t_R.

Table S3: **Likelihood for censored data.** For each type of censored data, the likelihood of that observation is a different function of f, depending also on the timing of the pertinent test(s).
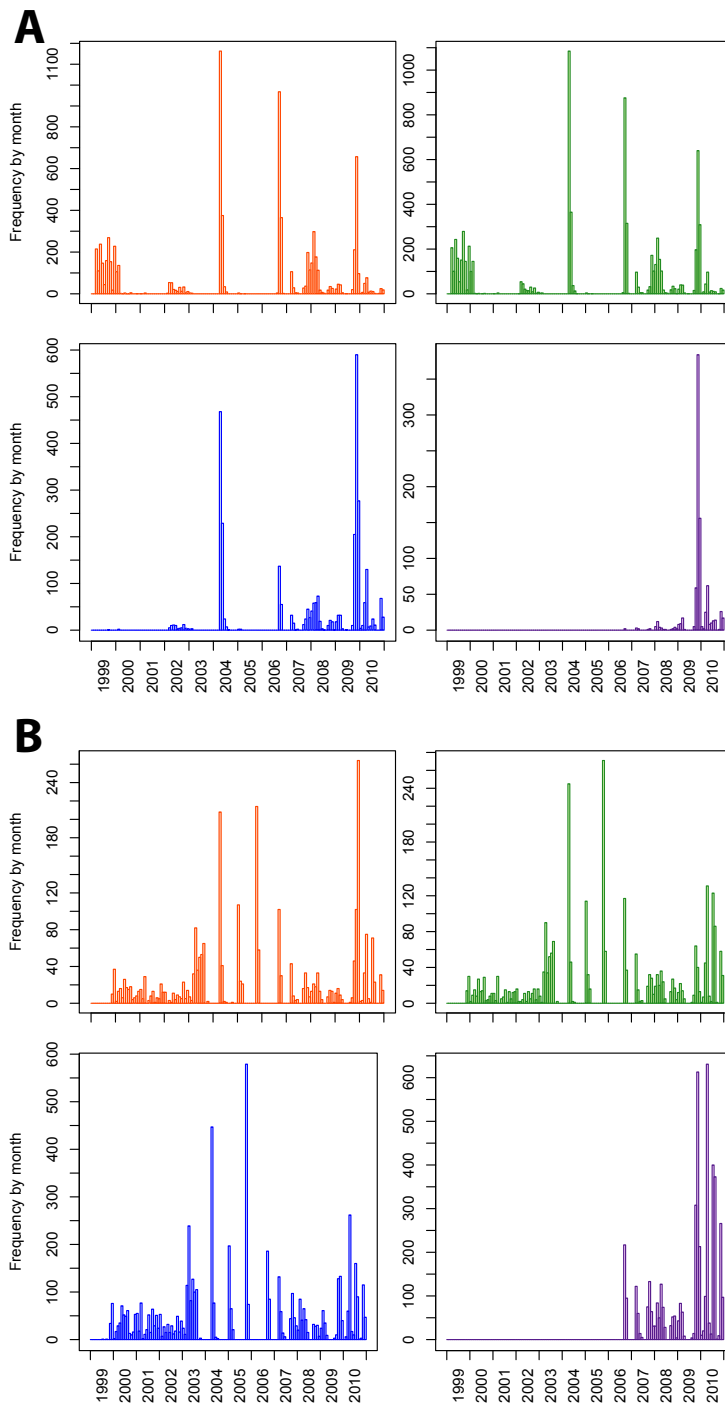
Fig. S1: **Left and right censored data.** (a) The number of left censored infections per month are plotted against time (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)). The three large spikes in the number of left censored infections correspond to the beginning of three of the cohort studies. (b) The number of right censored infections per month (i.e., the date of the final negative PRNT for an individual before they left the study) are plotted against time (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)). The DVCS study from April 2004 February 2006 consisted of three tests. Spikes in right censored data in mid-2004, early 2005 and late 2005 corresponded to individuals within that cohort who left after the first, second or third test respectively.

19
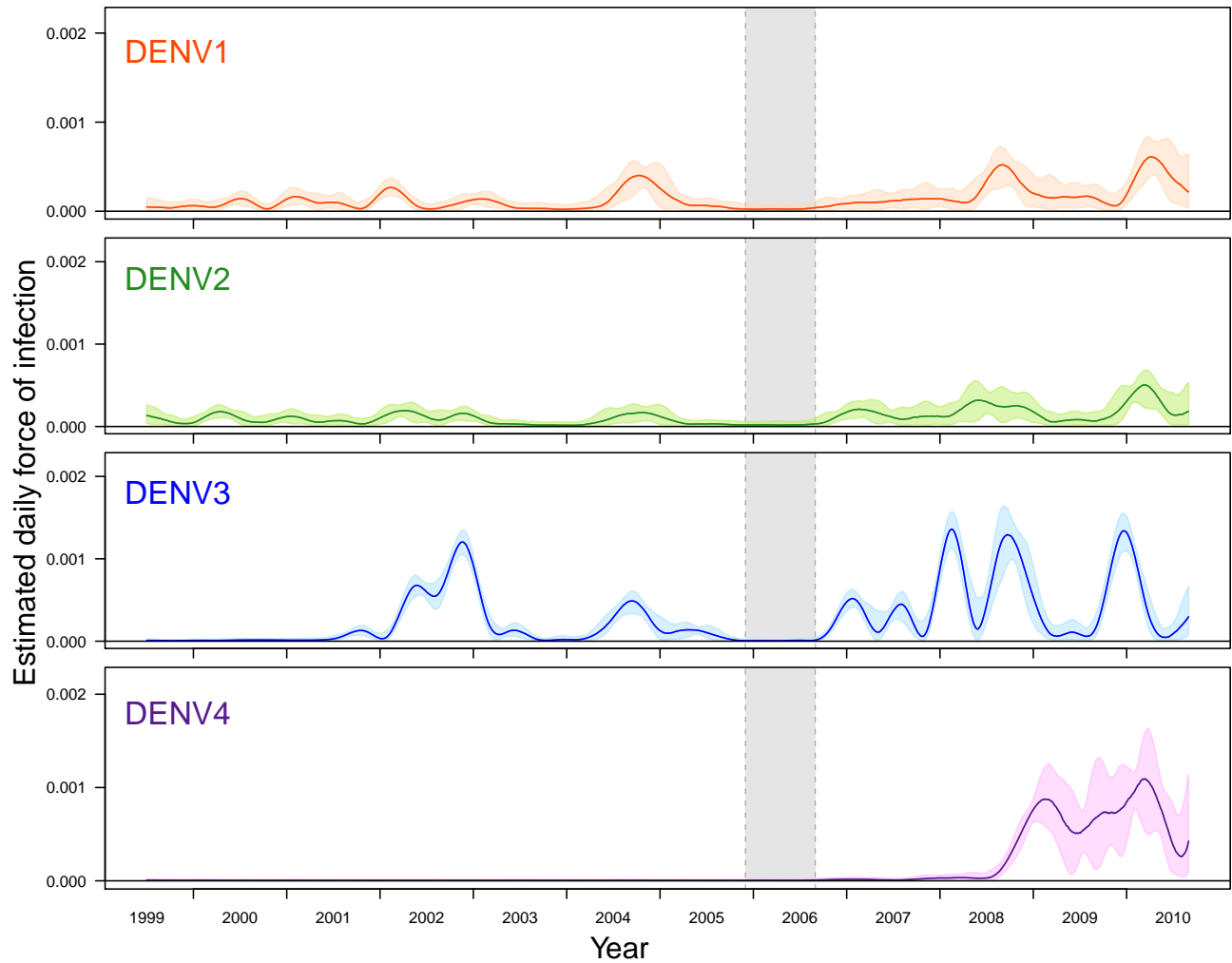
Fig. S2: **4 spline-per-year analysis: Daily estimates of the *FoI*.** For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of the *FoI* as well as the 90% BCI are plotted against time. The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region.
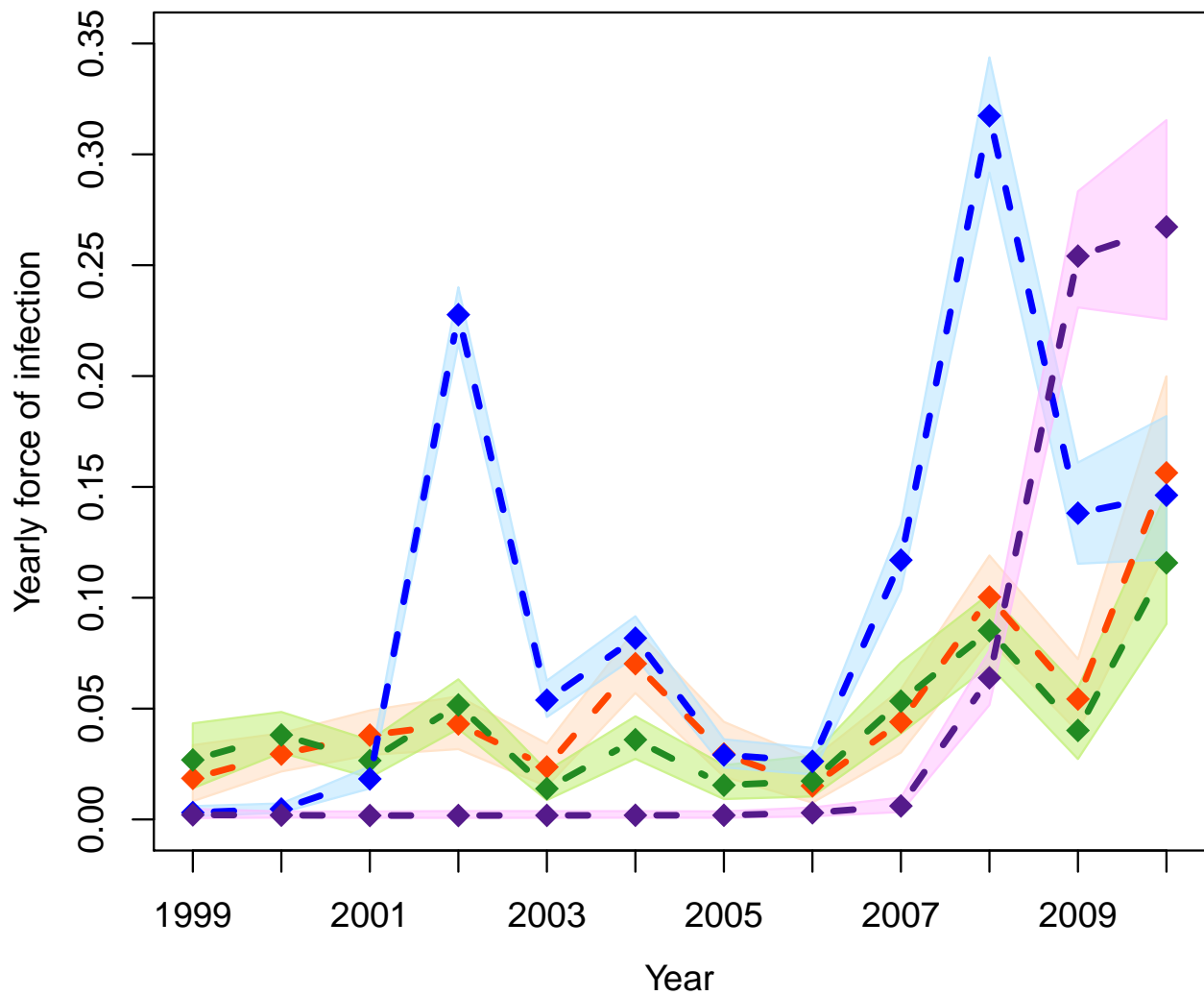
Fig. S3: **4 spline-per-year analysis: Yearly estimates of the *FoI*.** For each serotype (DENV-1 (orange), DENV-2 (green), DENV-3 (blue) and DENV-4 (purple)), yearly estimates of the *FoI* as well as the 90% BCI are plotted against time. The absence of a cohort study from late 2005 to mid-2006 does not preclude the estimation of the yearly *FoI* estimates for either 2005 or 2006 as evidenced by the non-zero *FoI* estimates for the circulating serotypes for both of those years.
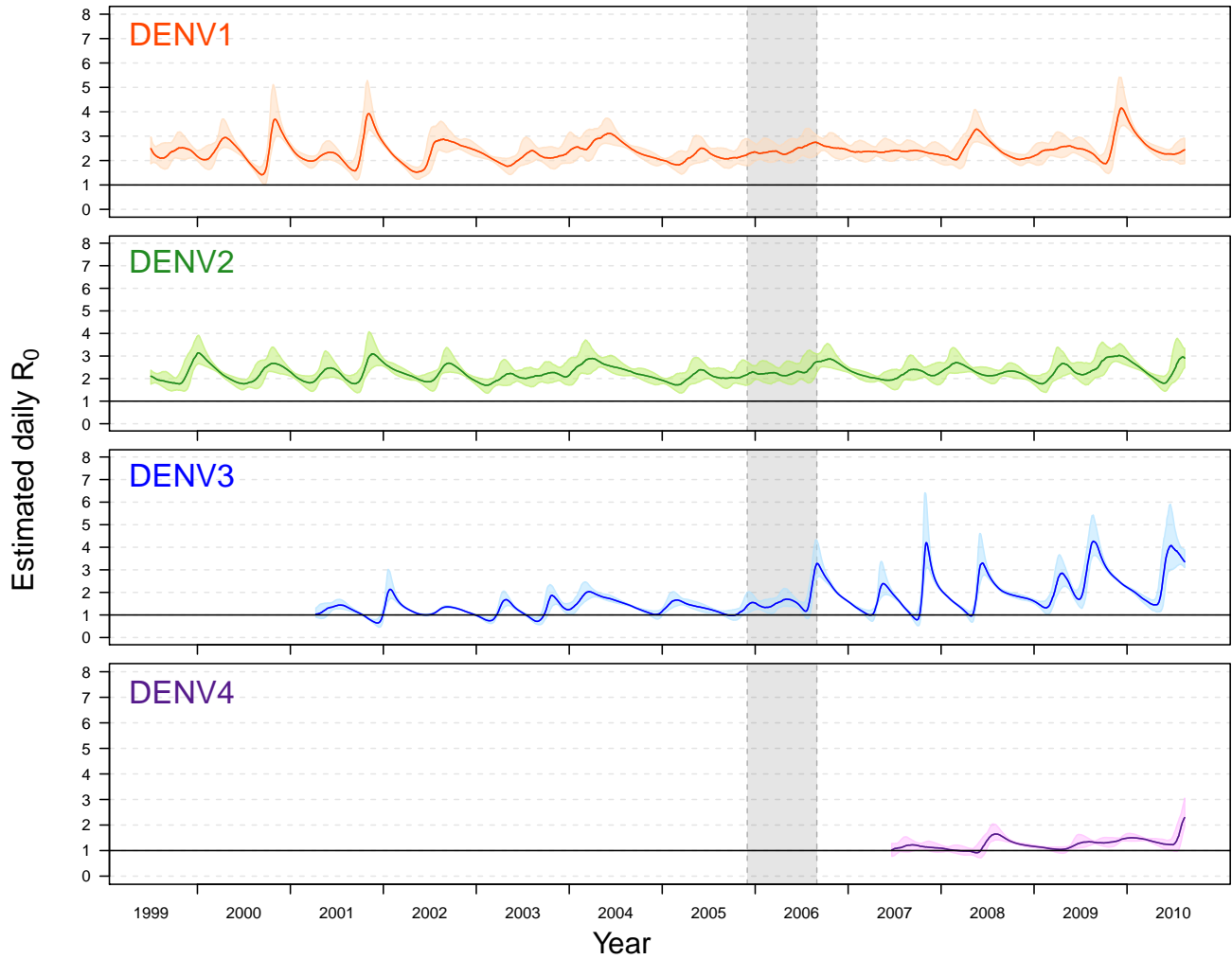
Fig. S4: **4 spline-per-year analysis: Daily estimates of** $\mathcal{R}_0$**.** For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of $\mathcal{R}_0$ as w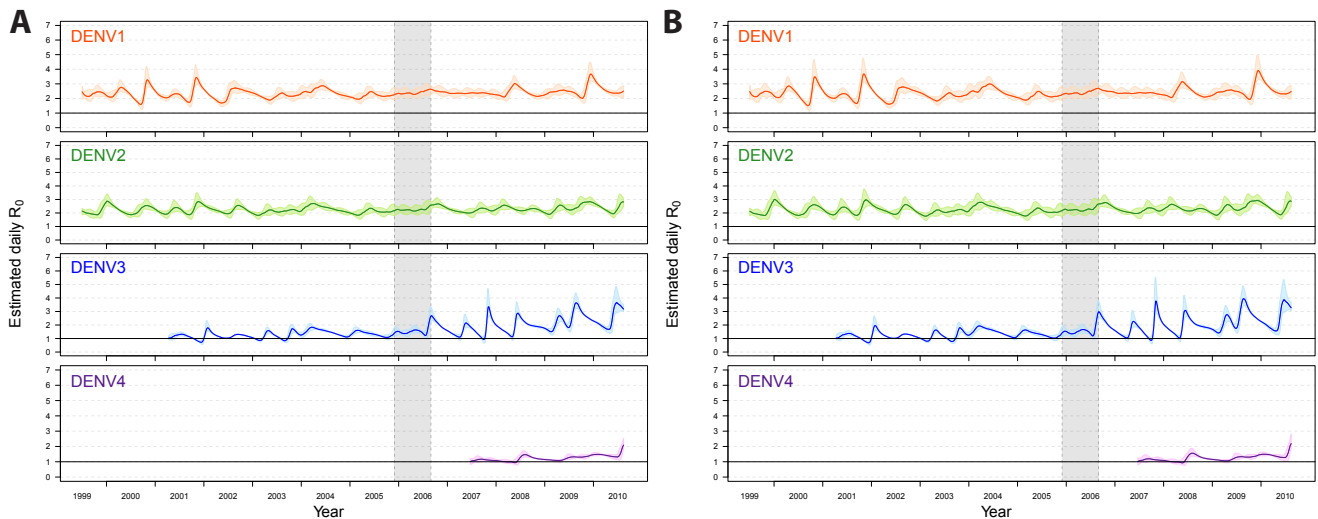ell as the 50% BCI are plotted against time. The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region. The estimates for both DENV-3 and DENV-4 are truncated, excluding estimation before their respective introductions.
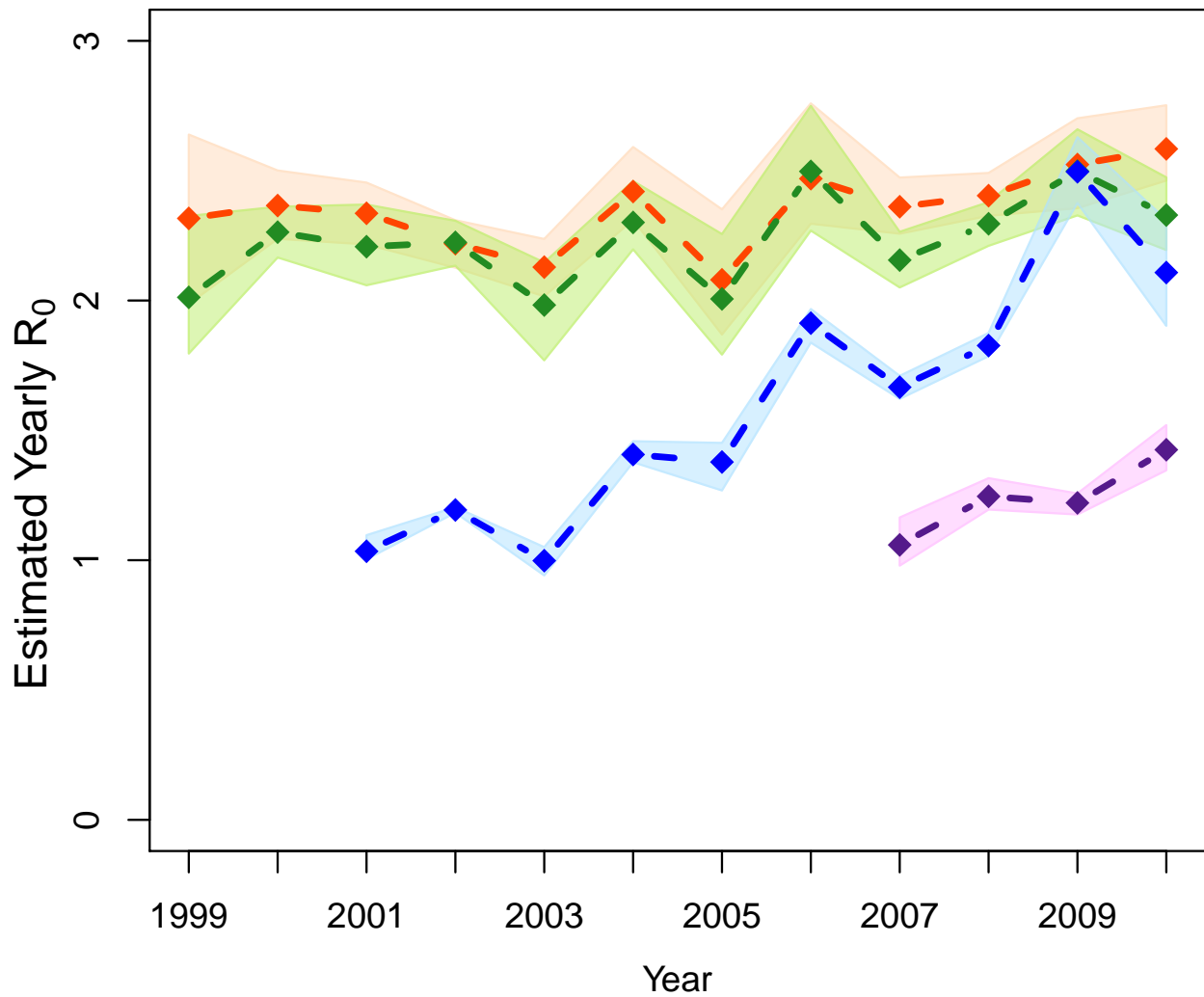
Fig. S5: **4 spline-per-year analysis: Yearly estimates of** $\mathcal{R}_0$**.** For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), yearly estimates of $\mathcal{R}_0$ as wel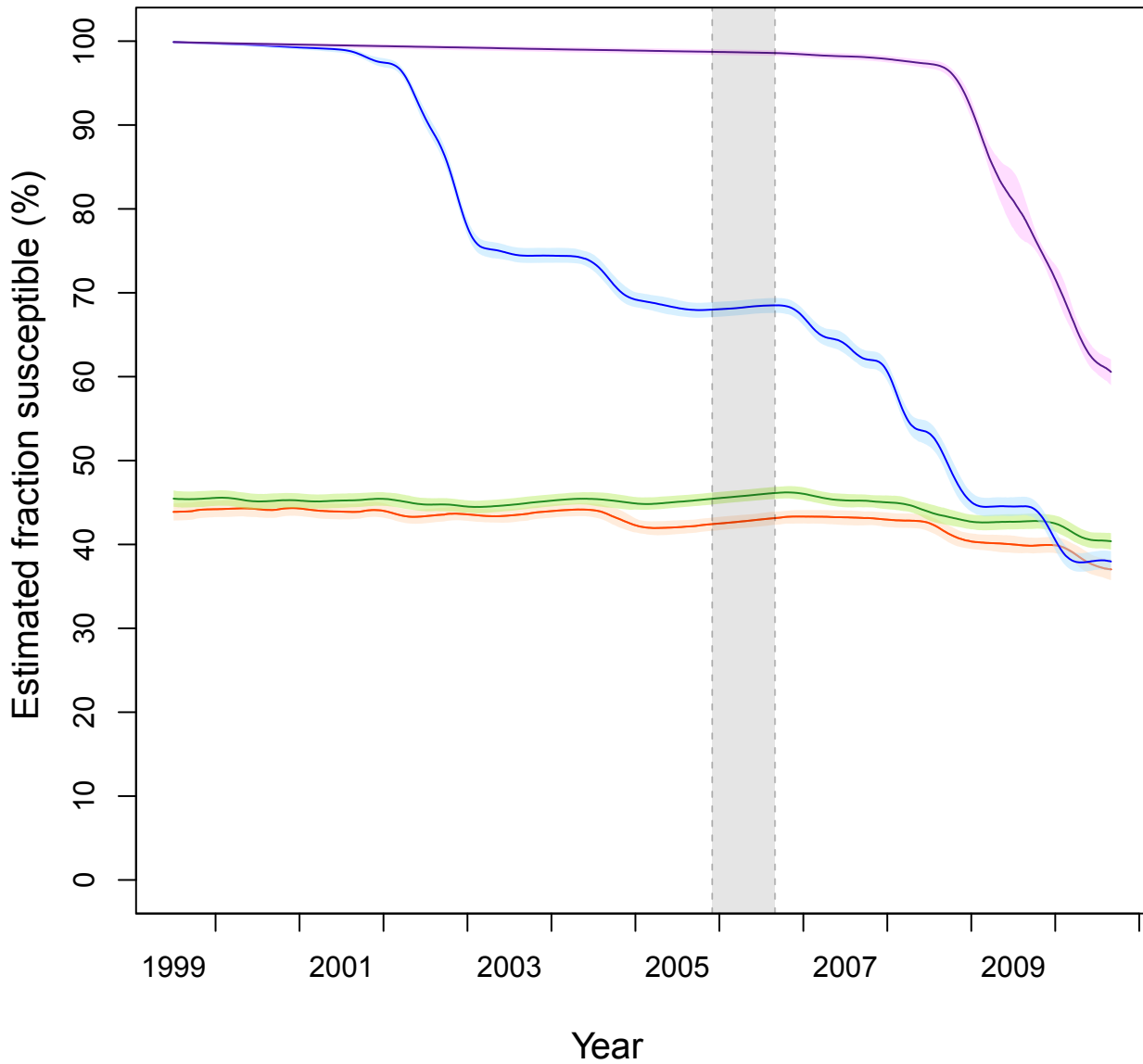l as the 90% BCI are plotted against time by computing the yearly mean of the daily estimates. The estimates for both DENV-3 and DENV-4 are truncated, excluding estimation before their respective introductions.
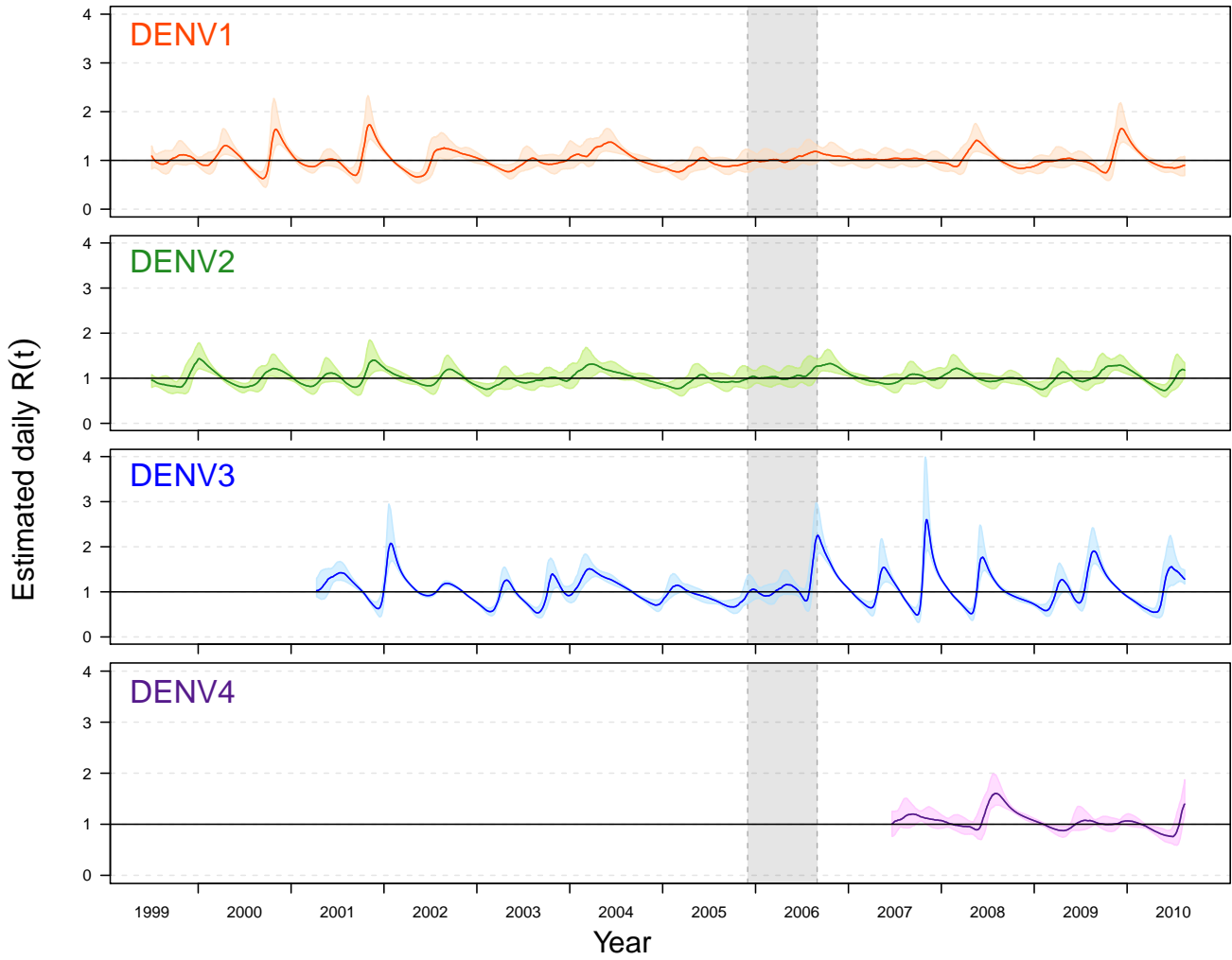
Fig. S6: **4 spline-per-year analysis: Daily estimates of the effective reproductive number,** $R(t)$. For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of $R(t)$ as well as the 50% BCI are plotted against time. The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region. The estimates for both DENV-3 and DENV-4 are truncated, excluding estimation before their respective introductions.
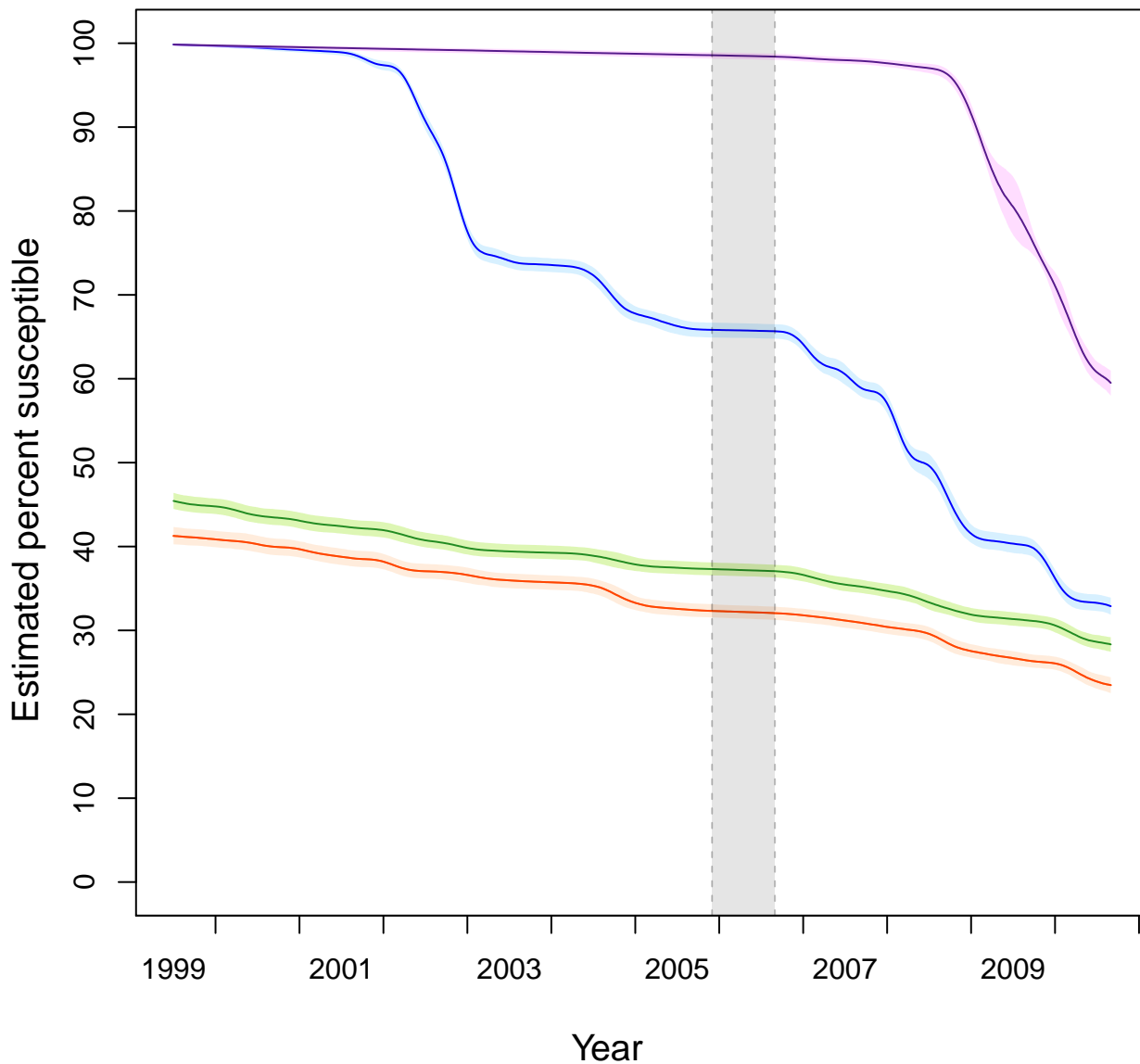
Fig. S7: **4 spline-per-year analysis: Fraction of the study population susceptible over time.** The fraction of the study population that was susceptible over time, $s(t)$, is plotted for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4. Note that because we define our study population to be those that were born before 1995 (and thus the susceptible pool was never replenished) and each serotype circulates after its introduction, these estimates decreased over time.
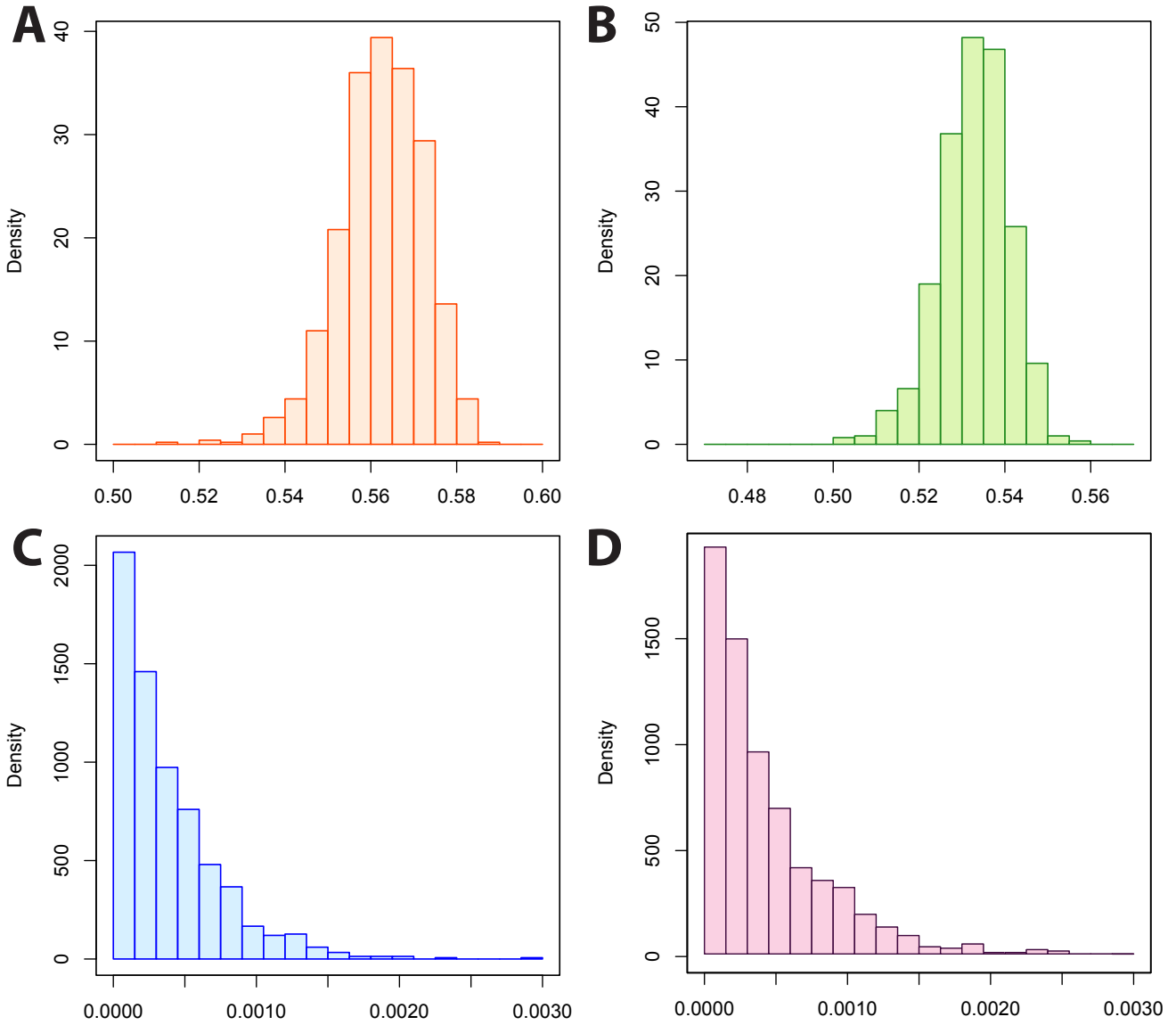
Fig. S8: **4 spline-per-year analysis: Posterior distribution of** κ**.** The posterior distribution of the fraction of the study population that was susceptible to each serotype at the time of the beginning of the study in 1999 is plotted for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4. Note that the scale is not the same for each of the figures.

Fig. S9: **4 spline-per-year analysis: Sensitivity of $\mathcal{R}_0$ to the serial interval, part 1.** We recomputed $\mathcal{R}_0$ using both a shorter and longer serial interval than found in [13]. In (a), we assumed that the length of time between primary and secondary infections was exactly 15 days. In (b), we lengthened the serial interval to 5 days, allowing the time between a primary and secondary infection to be between 15 and 19 days. In both (a) and (b), for each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of $\mathcal{R}_0$ as well as the 50% BCI are plotted against time.

Fig. S10: **4 spline-per-year analysis: Sensitivity of $\mathcal{R}_0$ to the serial interval, part 2.** Using the lower bound on both the EIP and the IIP for DENV reported in the literature, we estimated the shortest serial interval possible as 11 days. Using this value we recomputed $\mathcal{R}_0$. In (a), we assumed that the length of time between primary and secondary infections was between 11 and 13 days. In (b), we combined this lower bound (11 days) with the upper bound found in [13] (17 days), allowing the time between a primary and secondary infection to be between 11 and 17 days. In both (a) and (b), for each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of $\mathcal{R}_0$ as well as the 50% BCI are plotted against time.

Fig. S11: **4 spline-per-year analysis: Fraction of the entire population susceptible over time.** The fraction of the entire population that was susceptible over time shown as a percentage, $s_P(t)$, is plotted for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4.
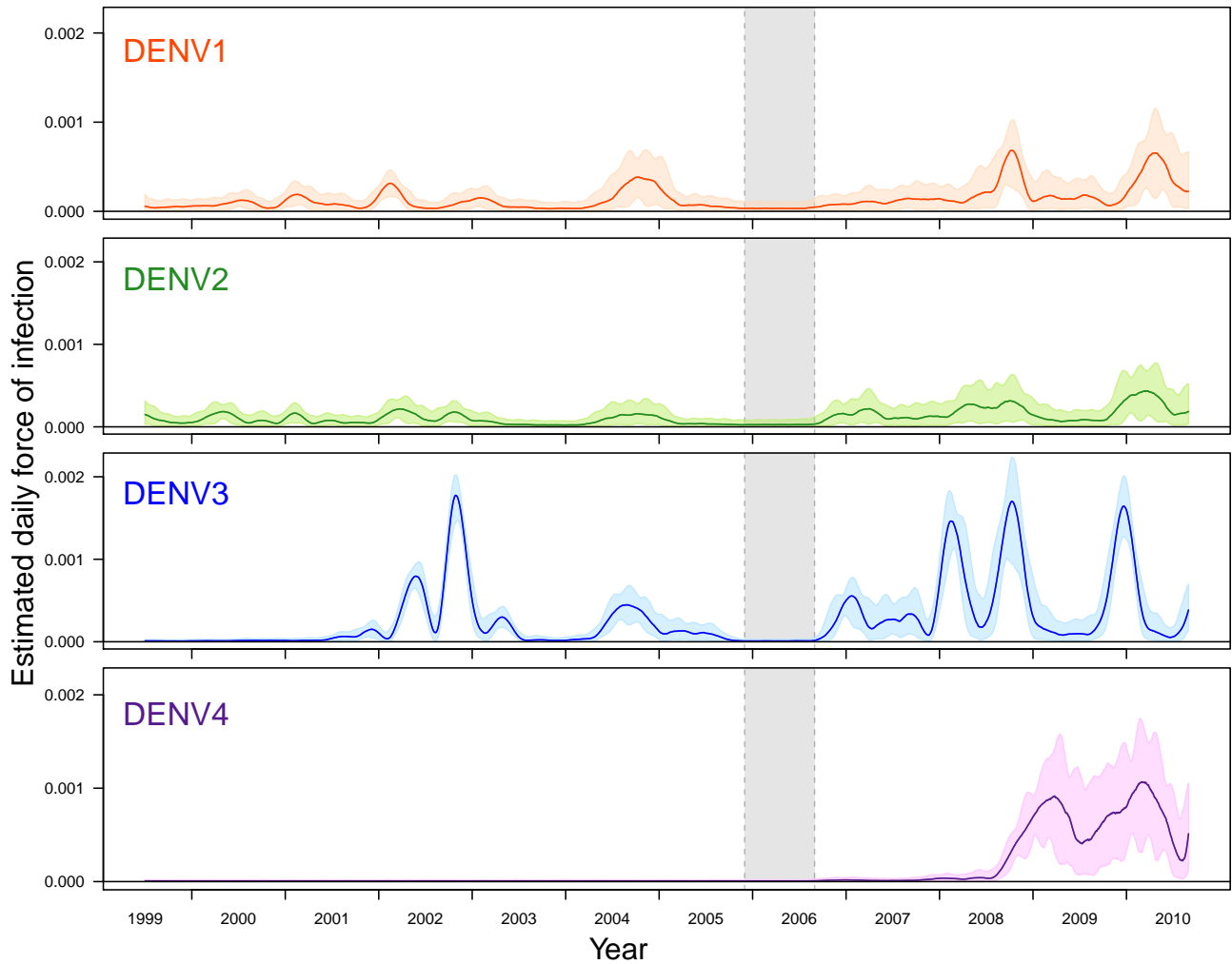
Fig. S12: **Sensitivity of daily estimates of the *FoI* to initial conditions.** Using different initial conditions for every parameter of every chain (uniformly drawing initial conditions from a uniform [0,0.001] distribution), *f* and then the daily *FoI* estimates were computed. For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of the *FoI* as well as the 90% BCI are plotted against time. The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region.
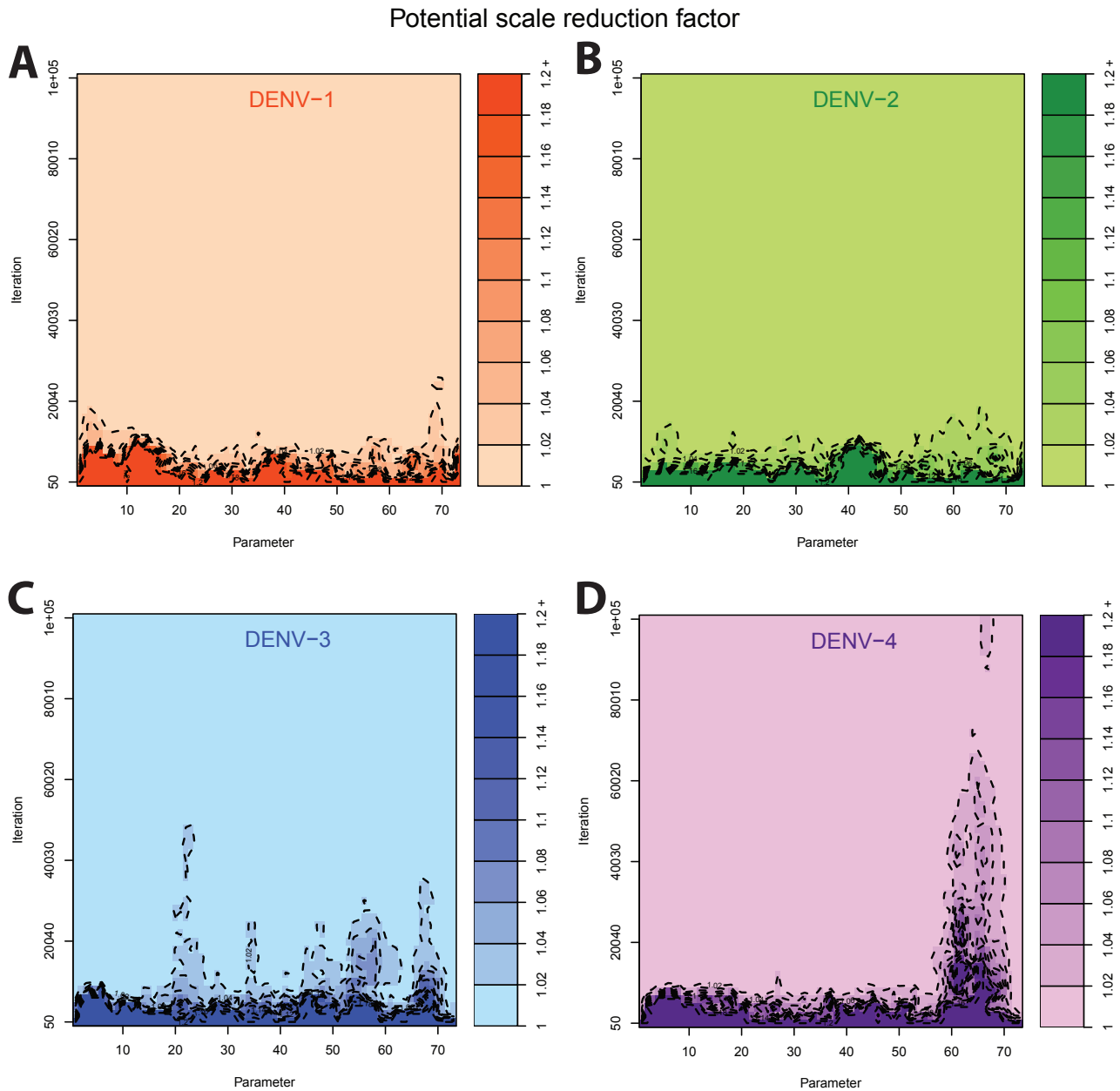
Fig. S13: **Potential scale reduction factors.** The potential scale reduction factors for computed for the MCMC chains corresponding to the parameters of (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4 are plotted against the iteration of the MCMC algorithm. A potential scale reduction greater than 1.1 indicated a lack of convergence for that particular parameter.
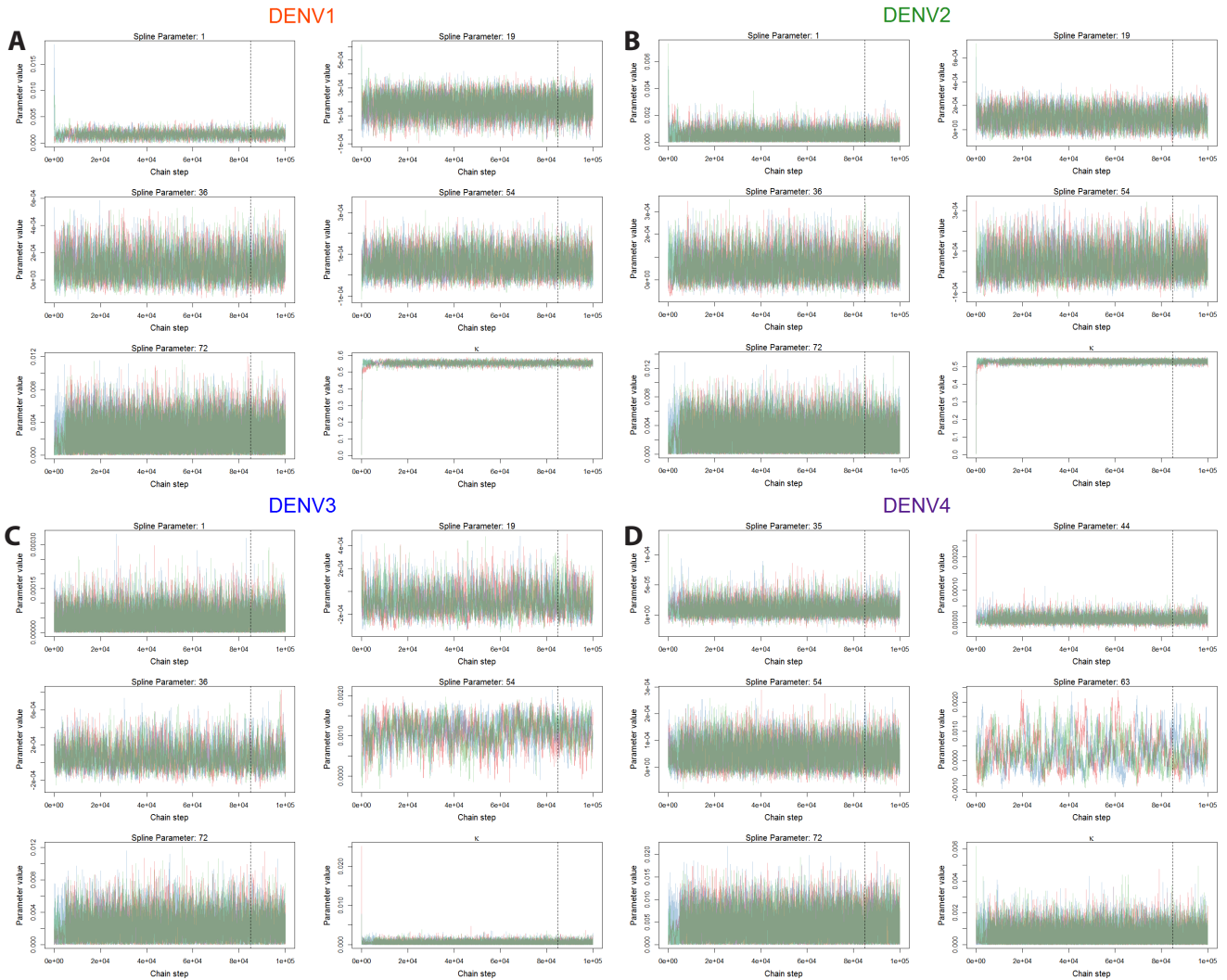
Fig. S14: **Trace plots.** Trace plots of a subset of parameters of the first three chains (red, blue and green lines respectively) are plotted for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4 against the iteration of the MCMC algorithm. The parameters selected for DENV-4 were altered due to the late invasion date of this serotype.
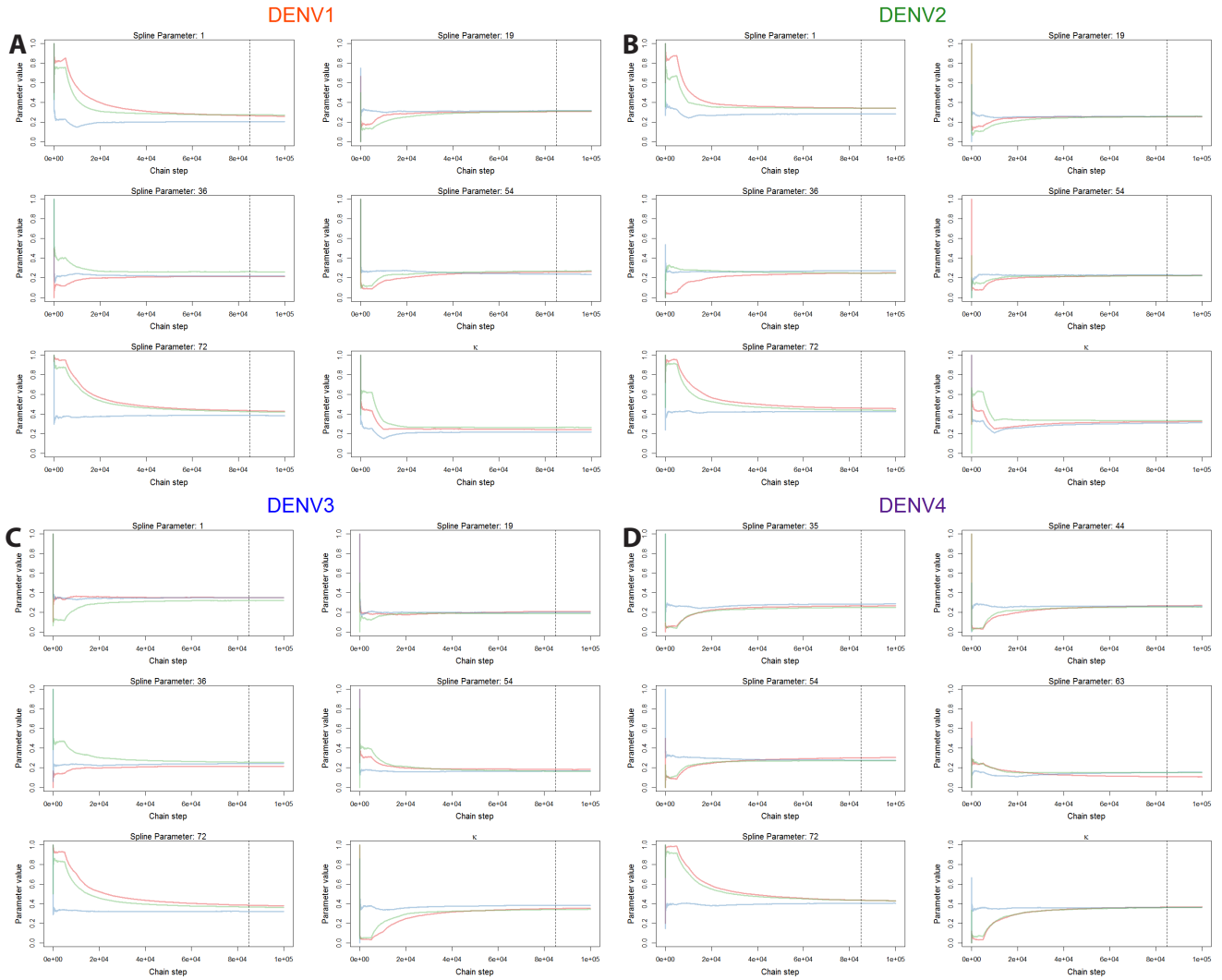
Fig. S15: **Estimates of the acceptance probabilities.** The estimated acceptance probabilities were running averages of the ratio of accepted proposals divided by the total number of values proposed. Estimated acceptance probabilities for a subset of parameters of the first three chains (red, blue and green lines respectively) are plotted for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4 against the iteration of the MCMC algorithm. The parameters selected for DENV-4 were altered due to the late invasion date of this serotype.
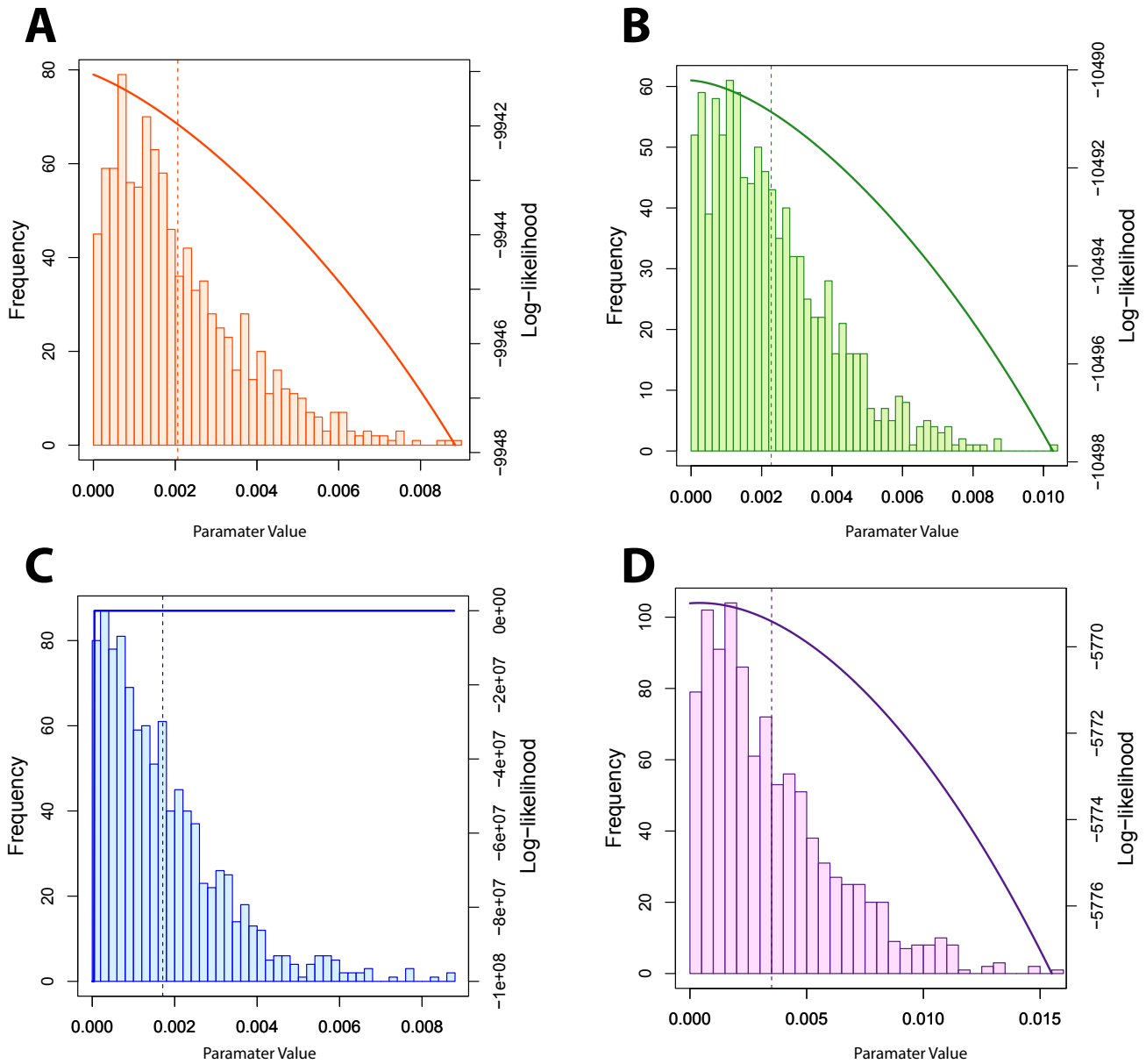
Fig. S16: **Posterior distribution and likelihood of final spline parameter.** The posterior distribution of the final spline parameter (histogram), posterior median of the final spline parameter (dashed line) and likelihood based on keeping all other parameters fixed at their posterior median and varying the final spline parameter (thick solid line) are plotted against values of the final spline parameter for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4. It is important to note that the solid line is not the profile likelihood, because the profile likelihood would be calculated by optimizing all other parameter values for any suggested value of the final spline parameter (and would thus decrease less sharply than the solid plotted line). In all cases however, it is clear that the same configuration of all other parameters and setting the final parameter equal to 0 (or a value much lower than the posterior median in the case of DENV-4) would have resulted in a better likelihood.
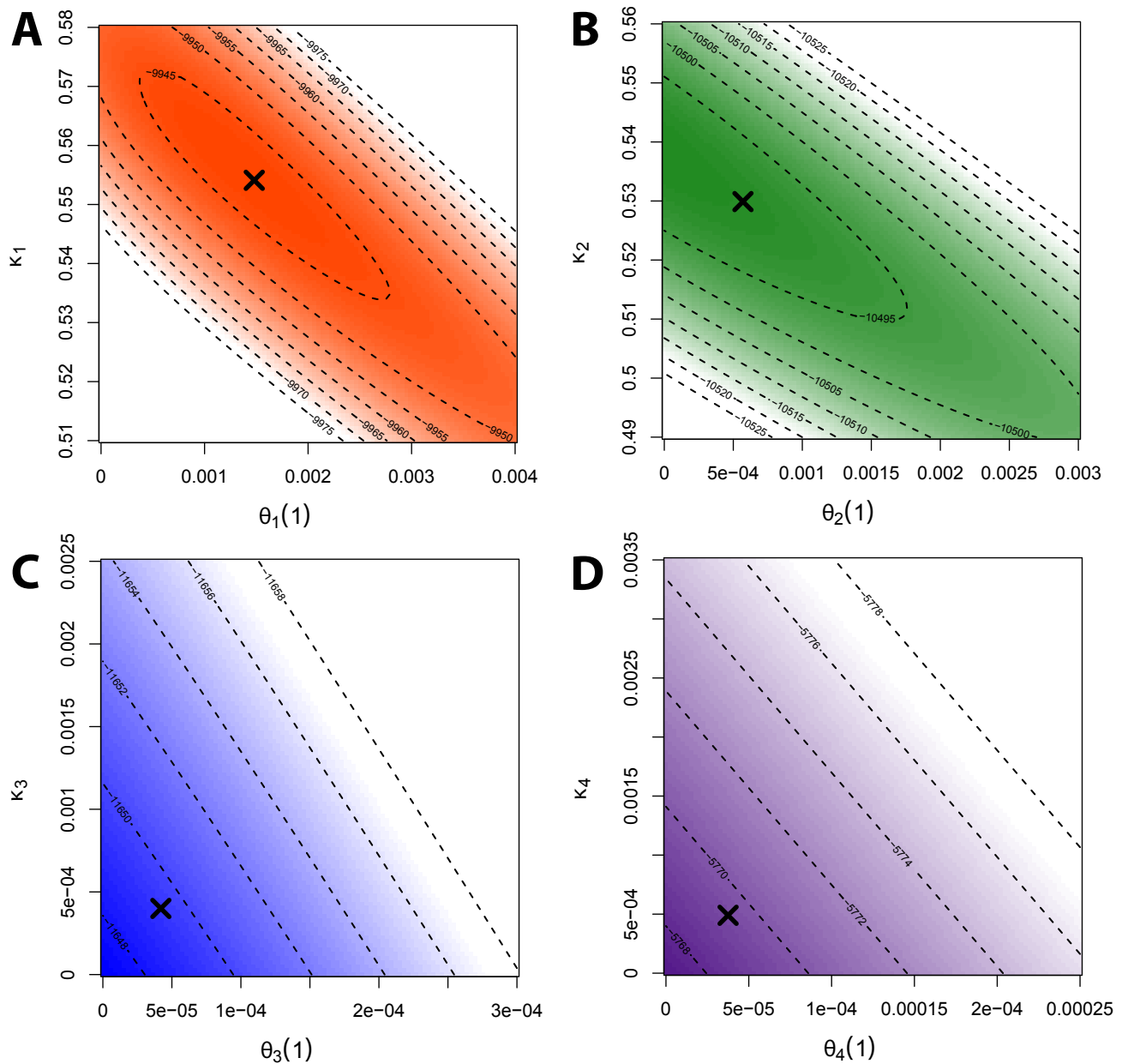
Fig. S17: **Identifiability issues between κ and the first spline parameter.** The 2-dimensional posterior distribution of κ and the first spline parameter are plotted for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4. In each case, there was a linear relationship between κ and the first spline parameter, especially for DENV-1 and DENV-2, making identification of the true value of the first spline parameter impossible in the current framework.
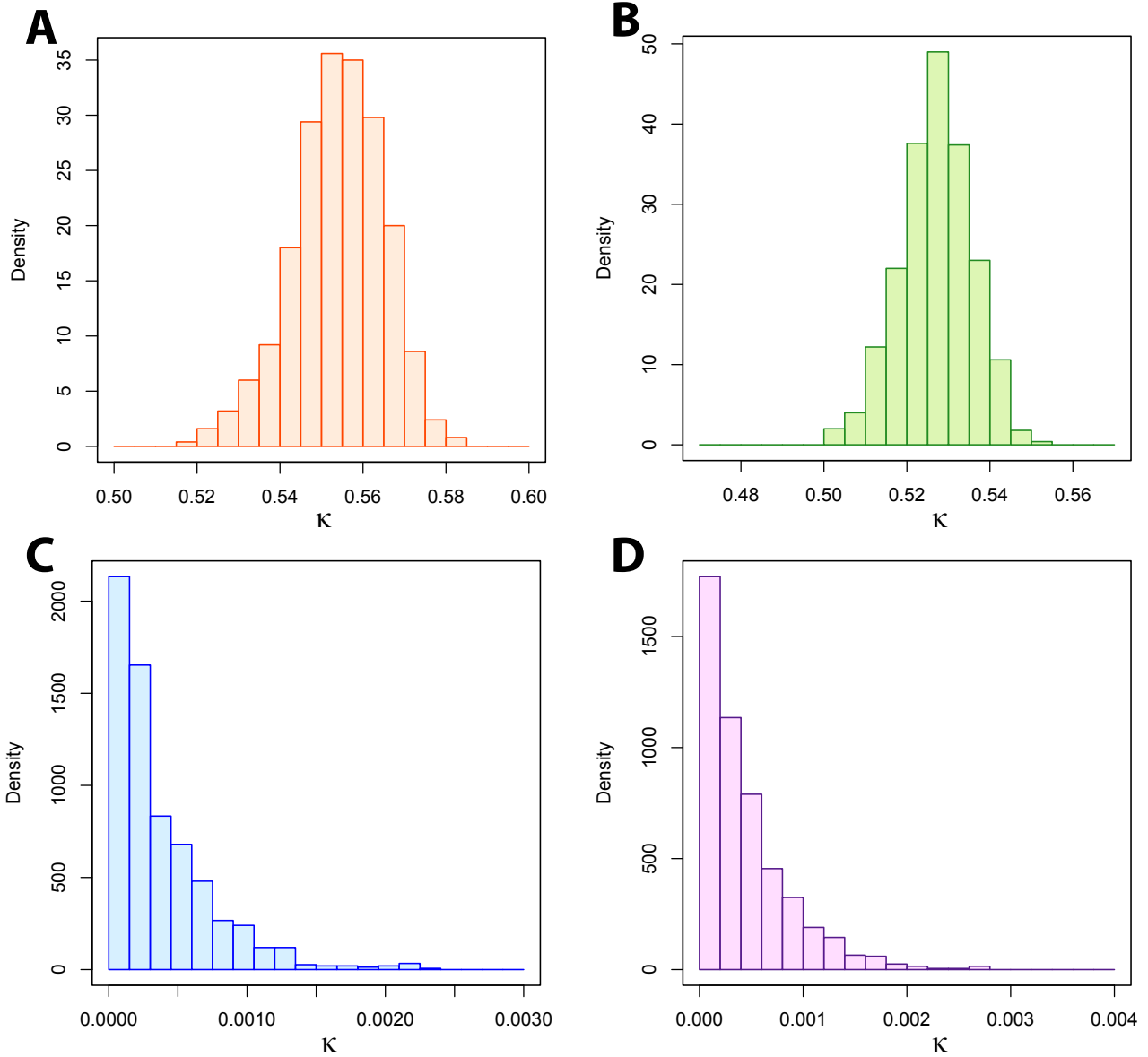
Fig. S18: **Posterior distribution of** κ. The posterior distribution of the fraction of the study population that was susceptible to each serotype at the time of the beginning of the study in 1999 is plotted for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4. Note that the scale is not the same for each of the figures.
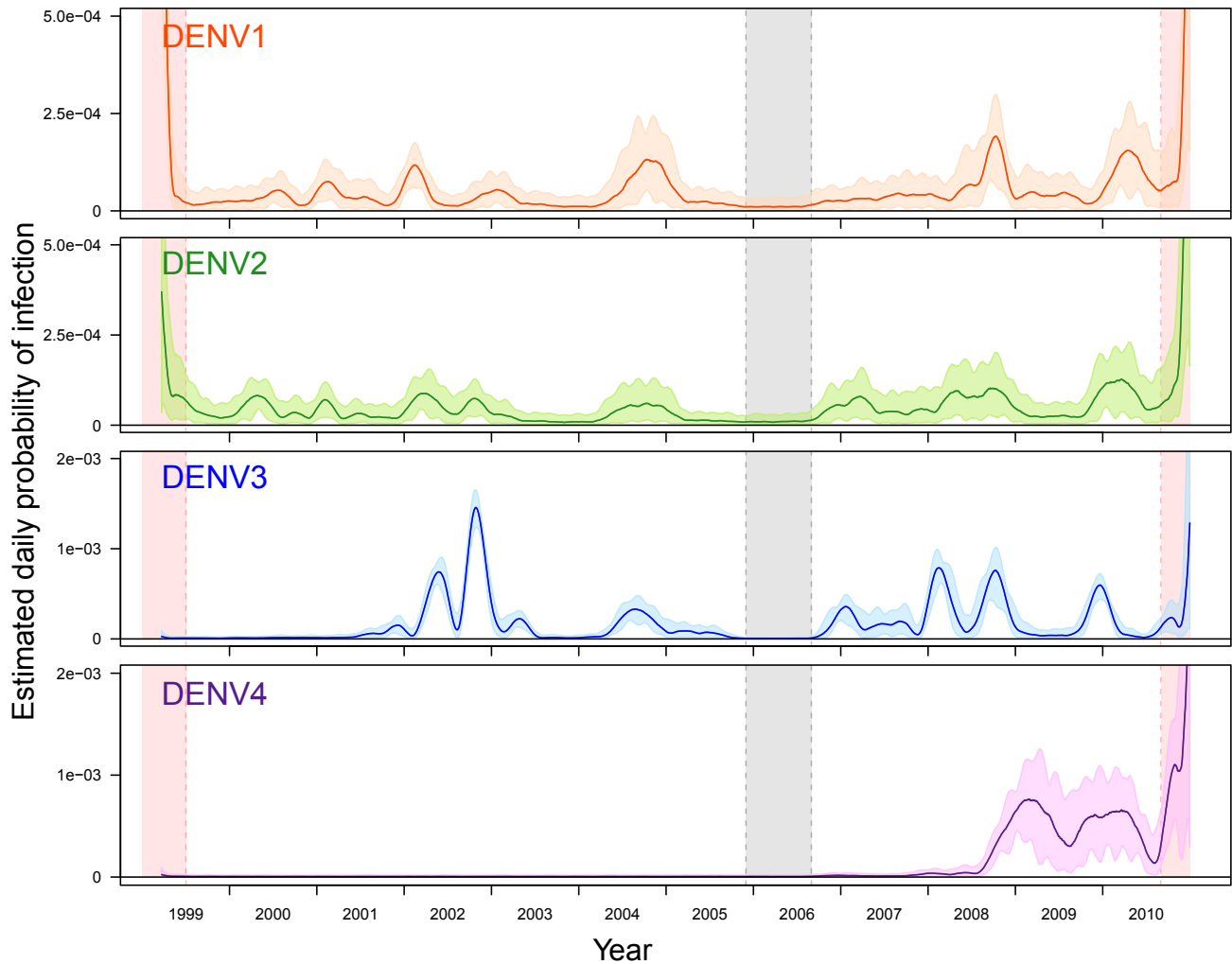
Fig. S19: **Estimated daily probability of infection, f.** For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of the probability of infection, f, as well as the 90% BCI are plotted against time. The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region. The red regions at the beginning and end of the plots indicate where the identifibility issues for the first and last spline parameter greatly affected the results.
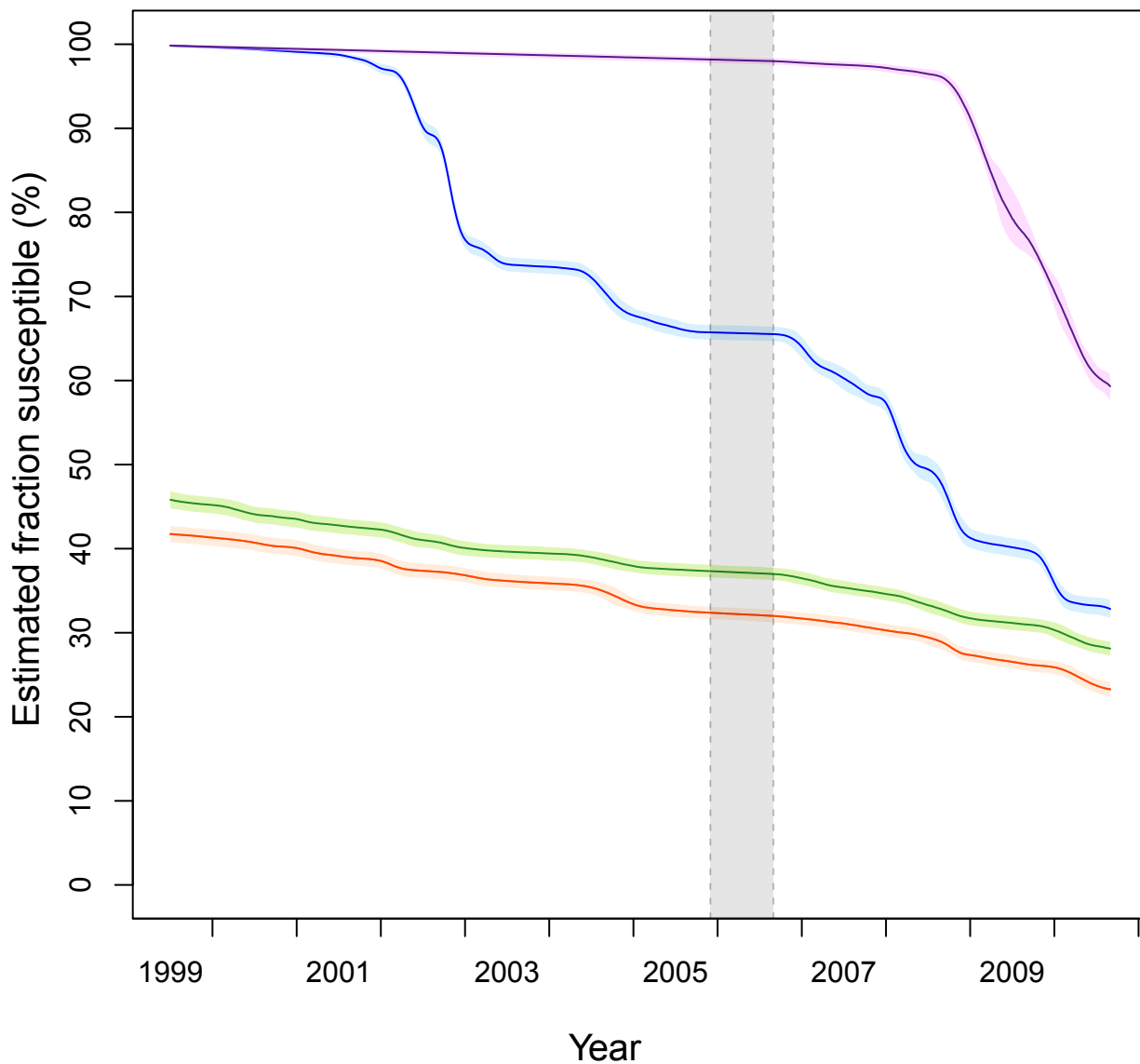
Fig. S20: **Fraction of the study population susceptible over time.** The fraction of the study population that was susceptible over time, $s(t)$, is plotted as a percentage for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4. Note that because we defined our study population to be those that were born before 1995 (and thus the susceptible pool was never replenished) and each serotype circulated after its introduction, these estimates decreased over time.
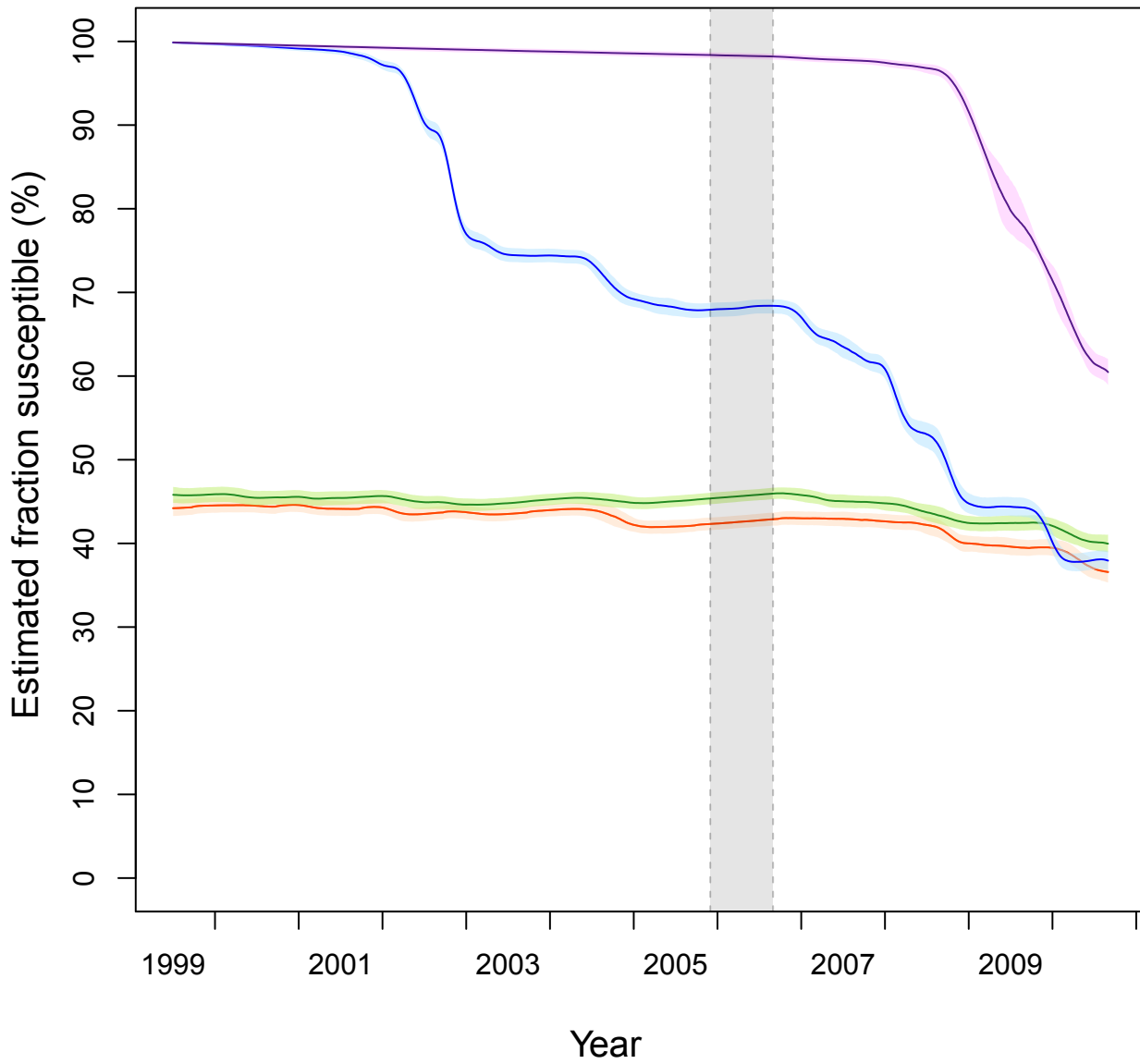
Fig. S21: **Fraction of the entire population susceptible over time.** The fraction of the entire population that was susceptible over time, $s_P(t)$, is plotted as a percentage for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4.
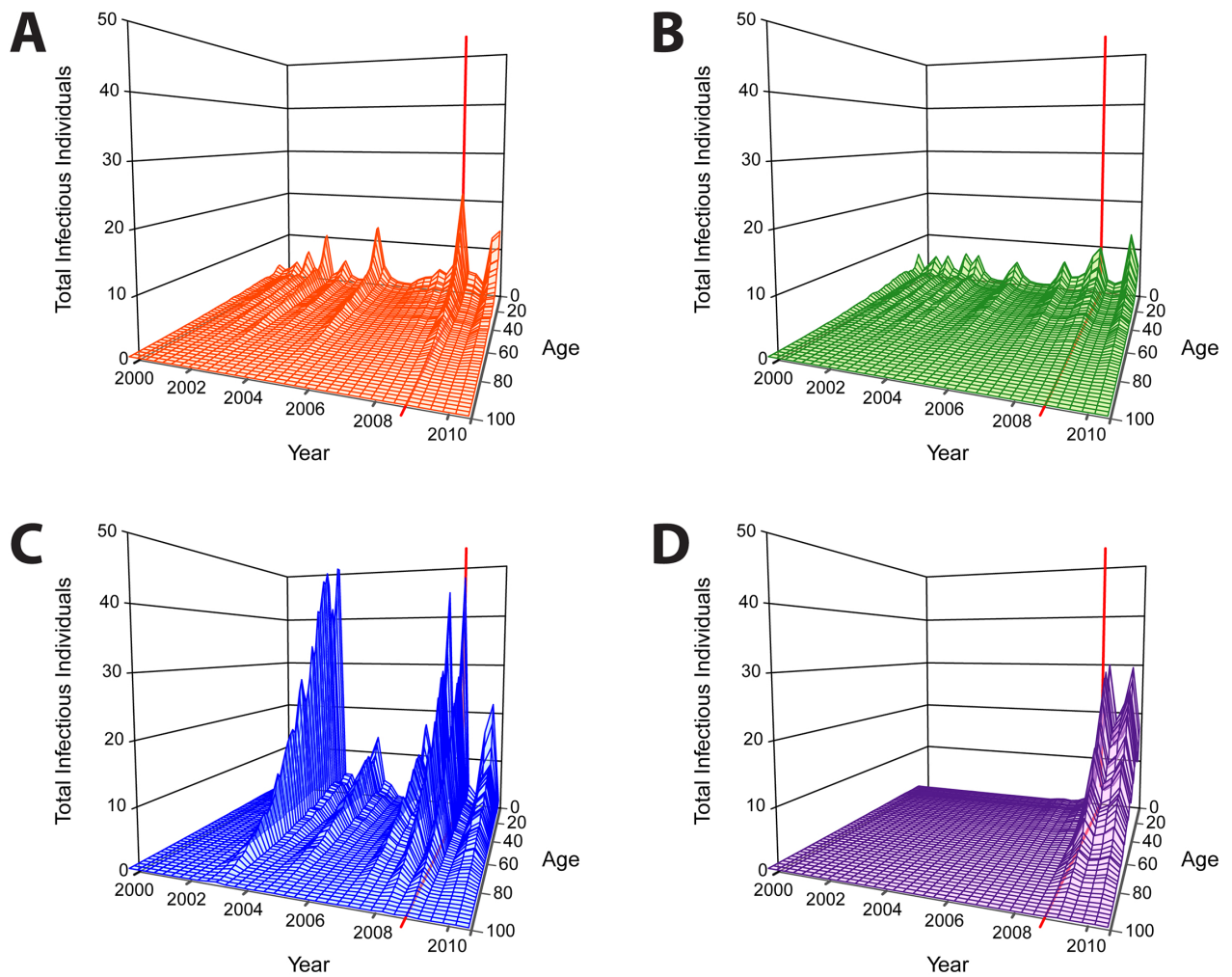
Fig. S22: **Estimated number of infectious individuals within Iquitos by age and day.** Using an age-structured SEIR model driven by the *FoI* estimates, the estimated number of individuals that are infectious on any given day are plotted by age for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4. The red line indicates October 1, 2008
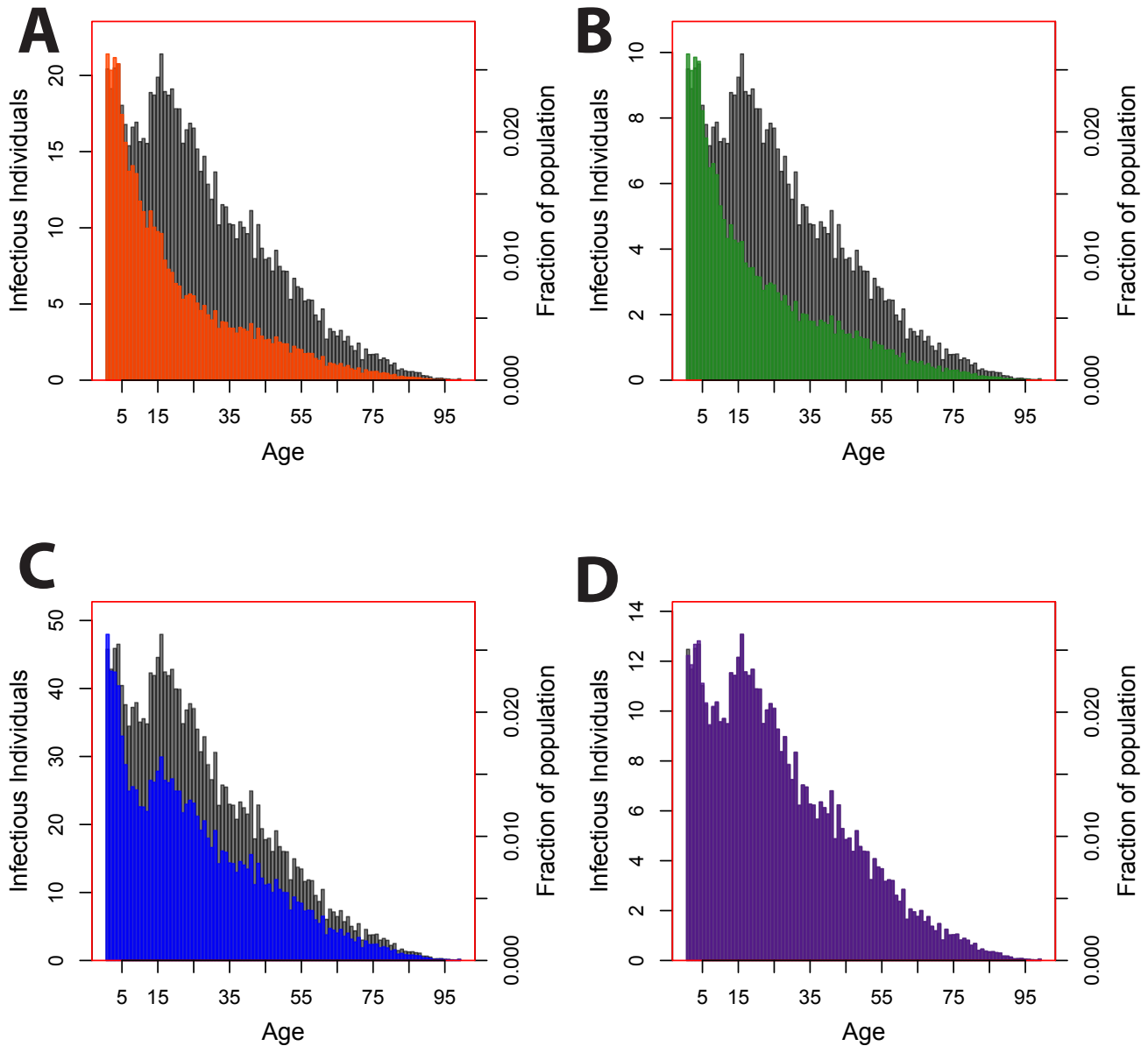
Fig. S23: **Predicted age distribution of infectious individuals within Iquitos.** The age distribution of infectious individuals on October 1, 2008, predicted from an age-structured SEIR model driven by the *FoI* estimates, is plotted for (a) DENV-1, (b) DENV-2, (c) DENV-3, and (d) DENV-4. The age distribution of the Iquitos population is plotted in dark gray.
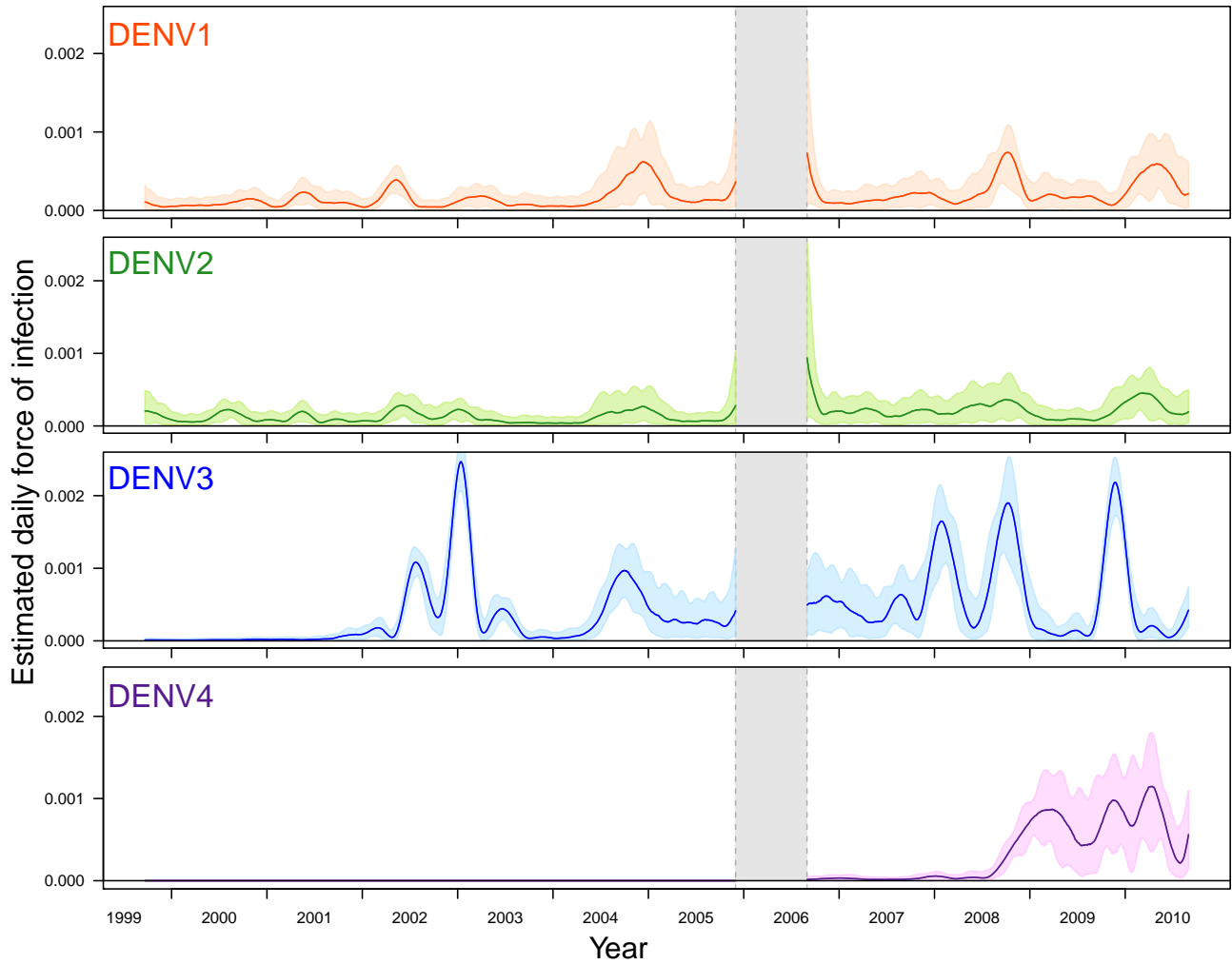
Fig. S24: **Robustness of daily estimates of the *FoI* to 2005-2006 gap.** Fitting two different models to the two halves of the data, using the gap from 2005-2006 to split the data, the resulting estimates of the *FoI* are plotted along with the 90% BCIs against time (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)). The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region.
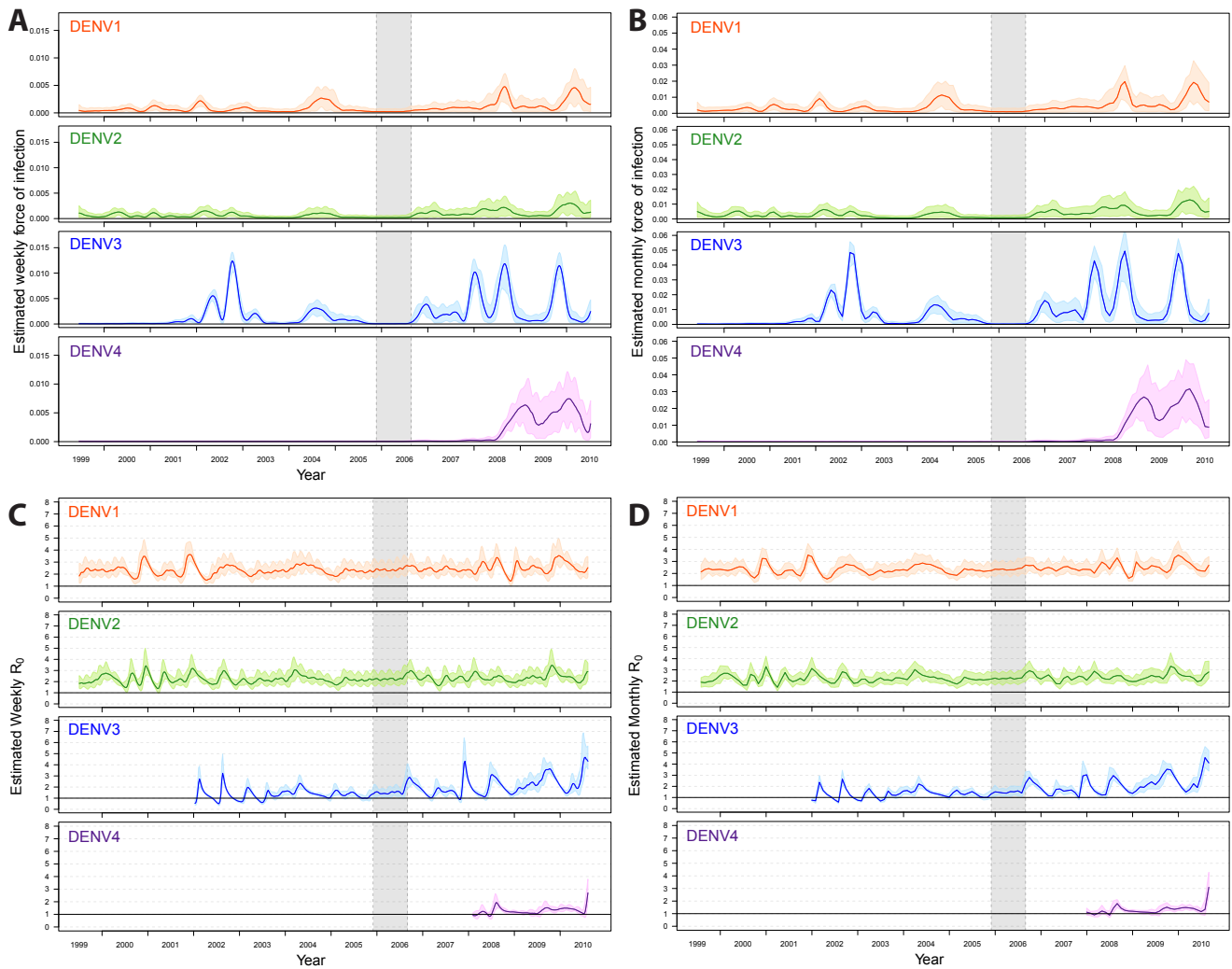
Fig. S25: **Weekly and monthly estimates of the *FoI* and $\mathcal{R}_0$.** In panel (a) and (b) respectively, weekly and monthly estimates of the *FoI* are plotted along with 90% BCIs. In panels (c) and (d) respectively, weekly and monthly averages of $\mathcal{R}_0$ are plotted along with 50% BCIs. For each panel, DENV-1 is plotted in the top sub-panel (orange), DENV-2 in the second sub-panel (green), DENV-3 in the third sub-panel (blue) and DENV-4 is plotted in the bottom sub-panel (purple)). The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region.
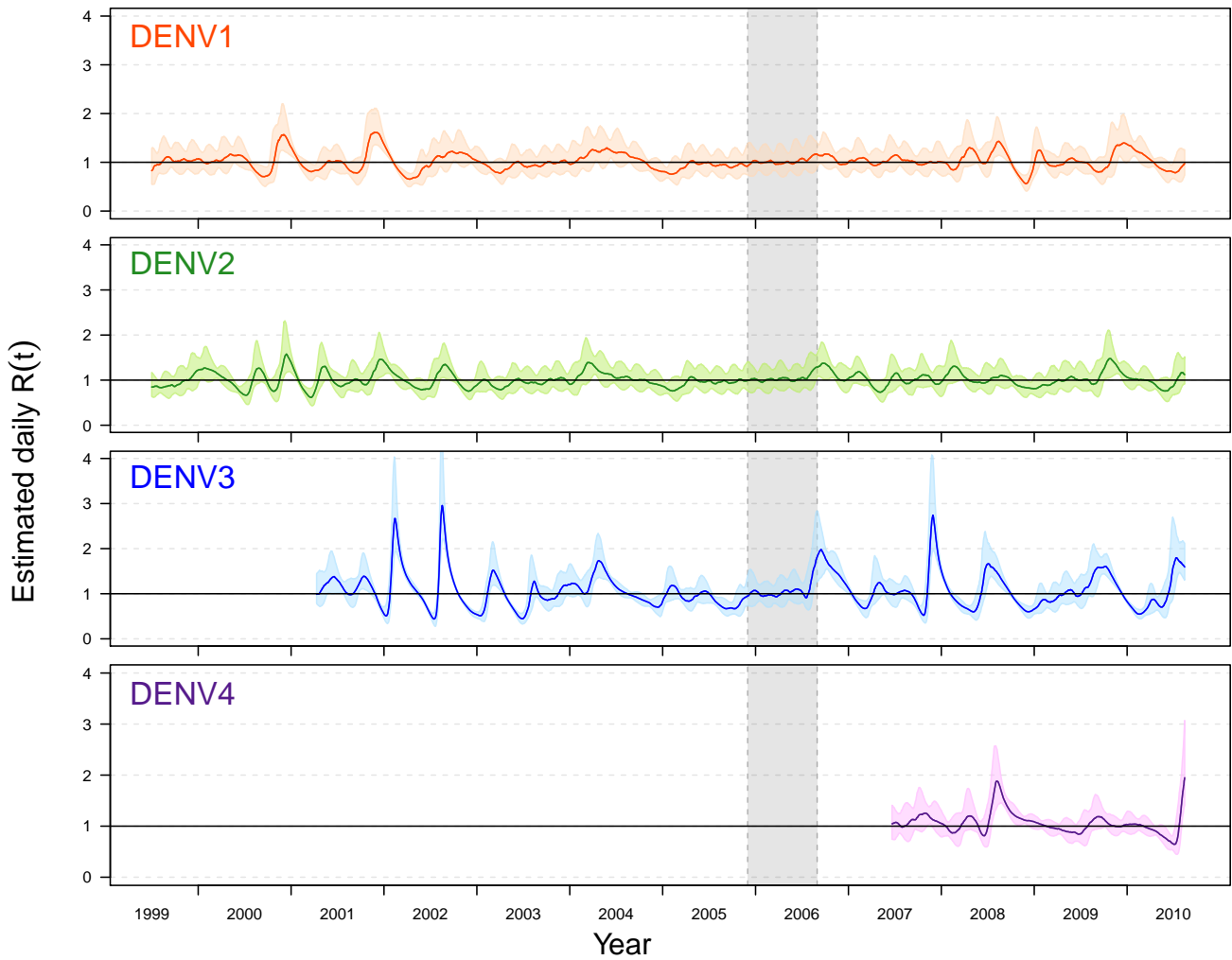
Fig. S26: **Daily estimates of the effective reproductive number,** $R(t)$. For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of R(t) as well as the 50% BCI are plotted against time. The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region. The estimates for both DENV-3 and DENV-4 are truncated, excluding estimation before their respective introductions.
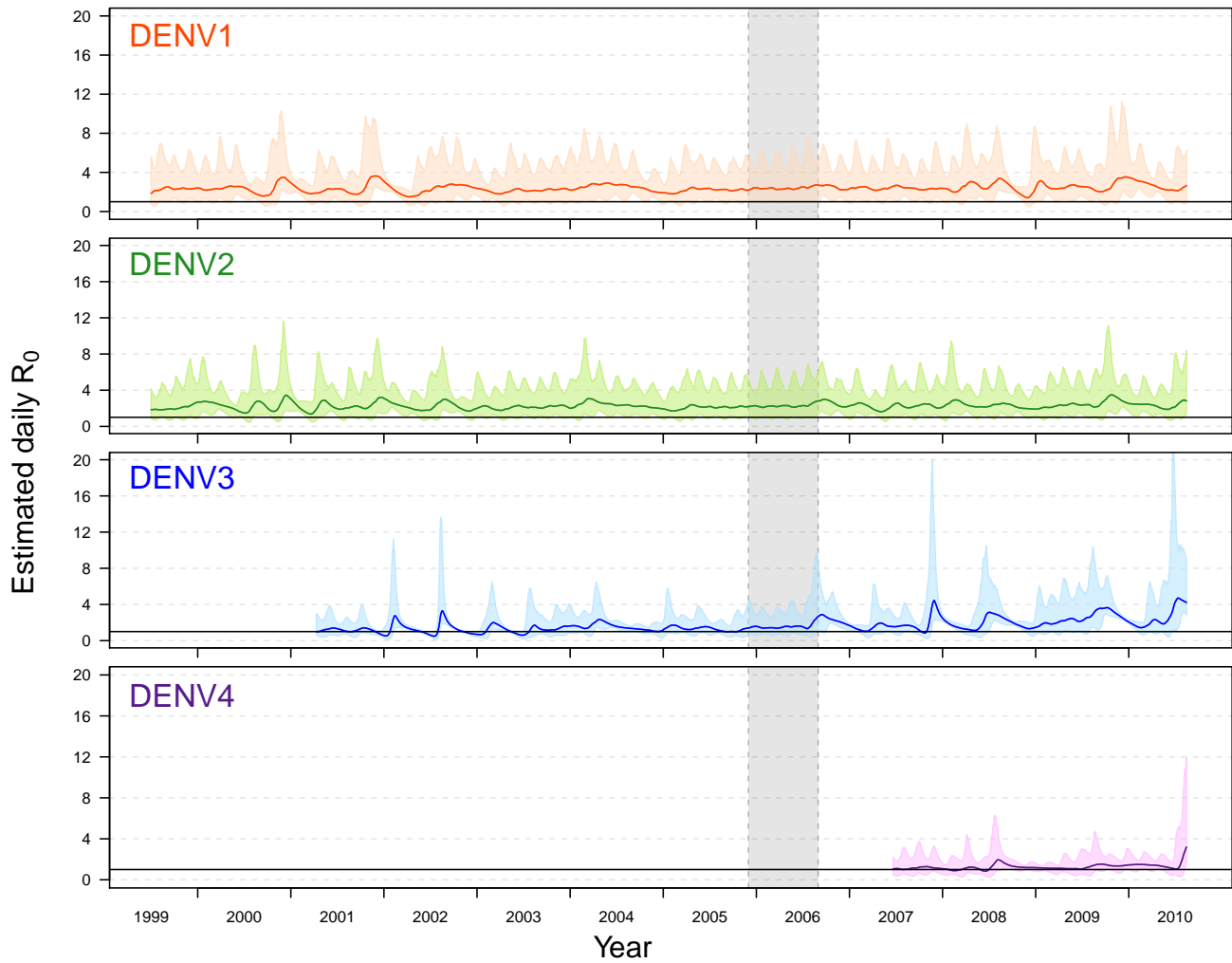
Fig. S27: **Daily estimates of** $\mathcal{R}_0$. For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of $\mathcal{R}_0$ as well as the 90% BCI are plotted against time. The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region. The estimates for both DENV-3 and DENV-4 are truncated, excluding estimation before their respective introductions. While the lower bound of the 90% BCI was comparable to that of the 50% BCI (Figure 5), the upper bound was considerably larger, in the extreme suggesting $\mathcal{R}_0$ values that exceeded 20 as being within the credible interval.
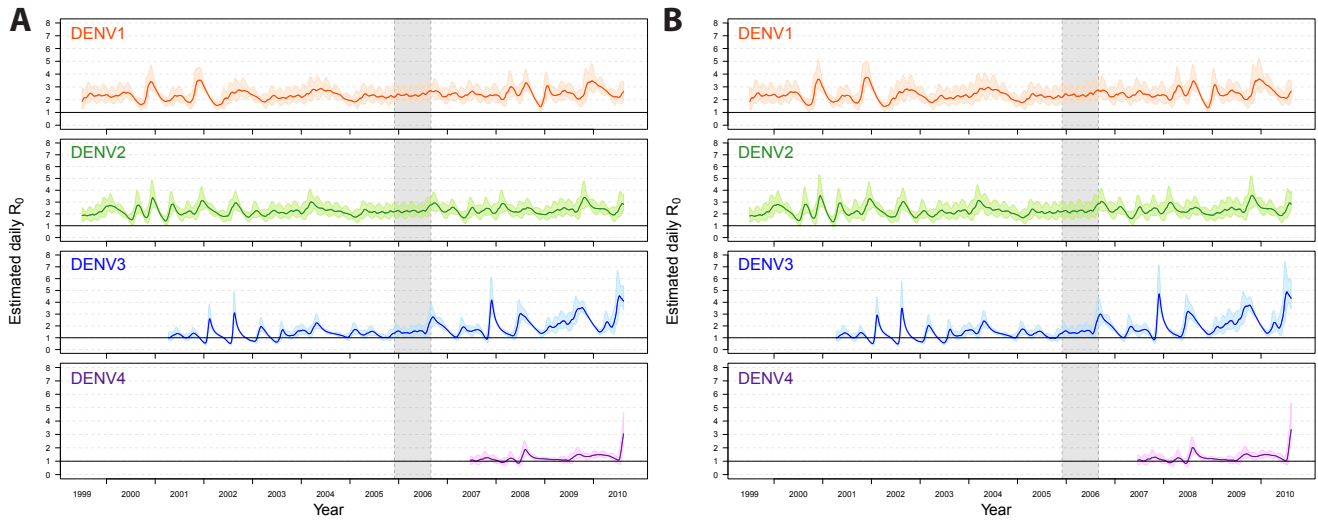
Fig. S28: **Sensitivity of $\mathcal{R}_0$ to the serial interval, part 1.** We recomputed $\mathcal{R}_0$ using both a shorter and longer serial interval than found in [13]. In (a), we assumed that the length of time between primary and secondary infections was exactly 15 days. In (b), we lengthened the serial interval to 5 days, allowing the time between a primary and secondary infection to be between 15 and 19 days. In both (a) and (b), for each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of $\mathcal{R}_0$ as well as the 50% BCI are plotted against time.
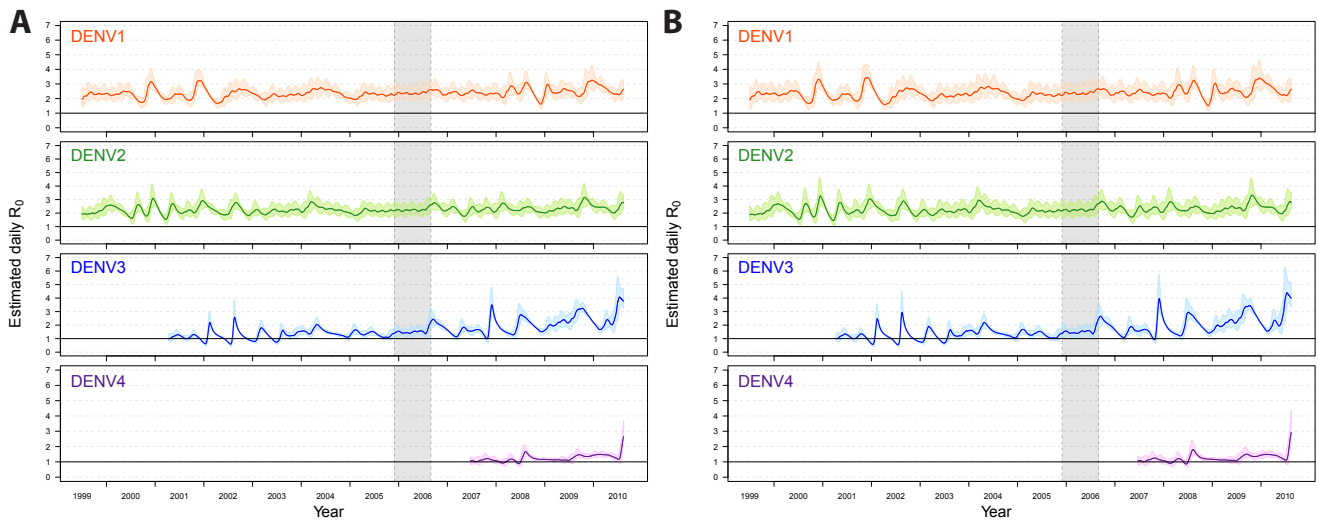

Fig. S29: **Sensitivity of $\mathcal{R}_0$ to the serial interval, part 2.** Using the lower bound on both the EIP and the IIP for DENV reported in the literature, we estimated the shortest serial interval possible as 11 days. Using this value we recomputed $\mathcal{R}_0$. In (a), we assumed that the length of time between primary and secondary infections was between 11 and 13 days. In (b), we combined this lower bound (11 days) with the upper bound found in [13] (17 days), allowing the time between a primary and secondary infection to be between 11 and 17 days. In both (a) and (b), for each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of $\mathcal{R}_0$ as well as the 50% BCI are plotted against time.
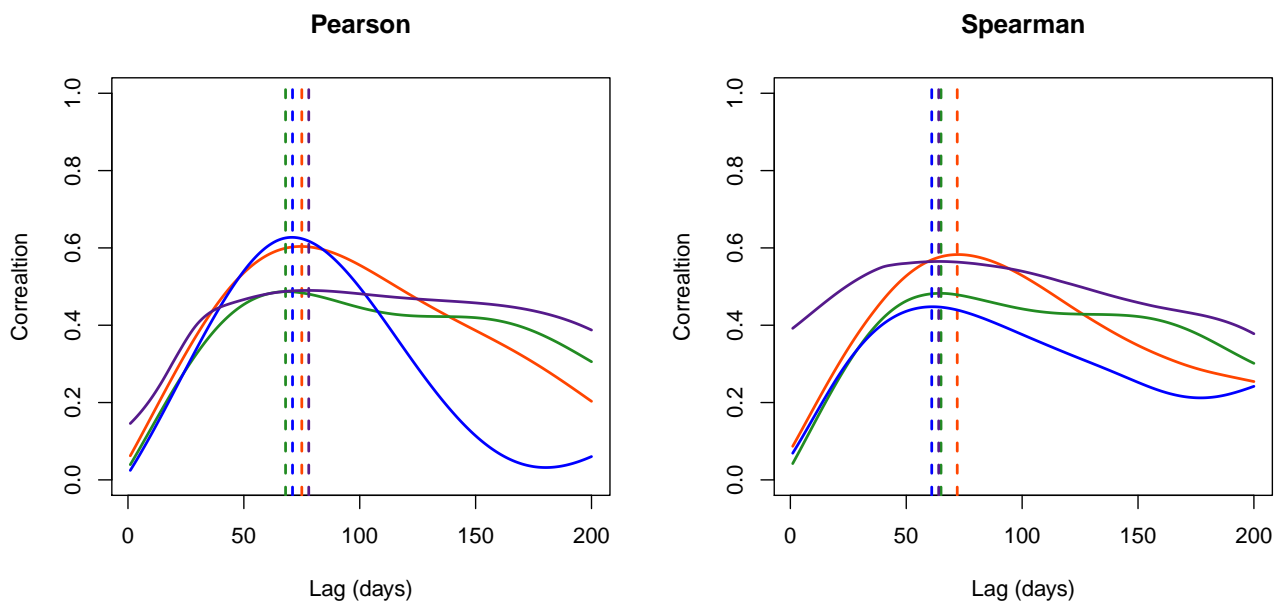
Fig. S30: **Lagged relationship between the *FoI* and $\mathcal{R}_0$.** The Pearson and Spearman rank correlations (left and right panel respectively) are plotted for various lags between the *FoI* and $\mathcal{R}_0$ for each serotype (DENV-1 (orange), DENV-2 (green), DENV-3 (blue) and DENV-4 (purple)). The maximum lags (illustrated by vertical dashed lines) were comparable for each serotype and were between 55 and 75 days.
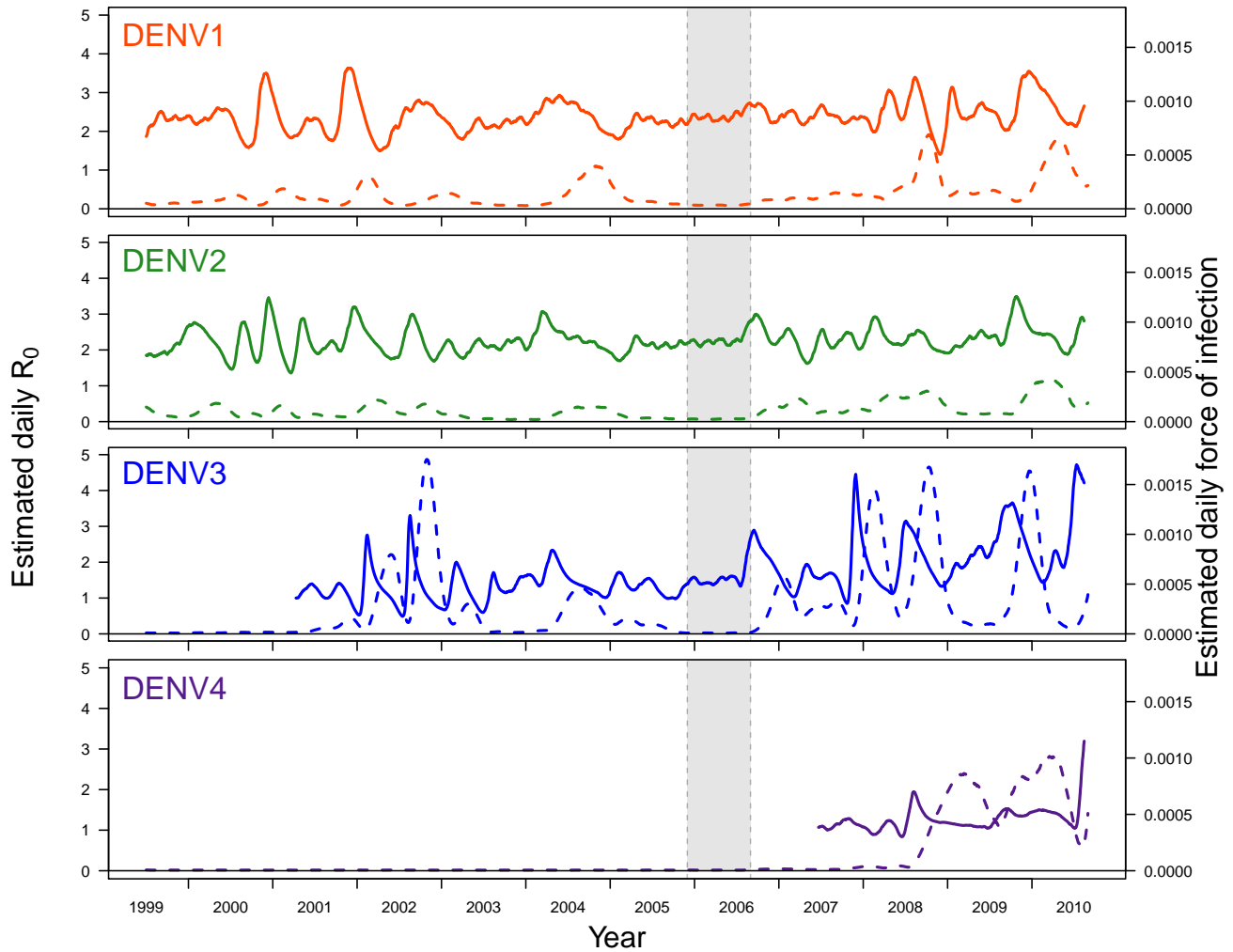
Fig. S31: **Relationship between the *FoI* and $\mathcal{R}_0$.** For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of the *FoI* (dashed lines) and $\mathcal{R}_0$ (solid lines) are plotted each other against time. The absence of a cohort study from late 2005 to mid-2006 is indicated by the grey shaded region. Every increase in the *FoI* was preceded by a sharp spike in $\mathcal{R}_0$.
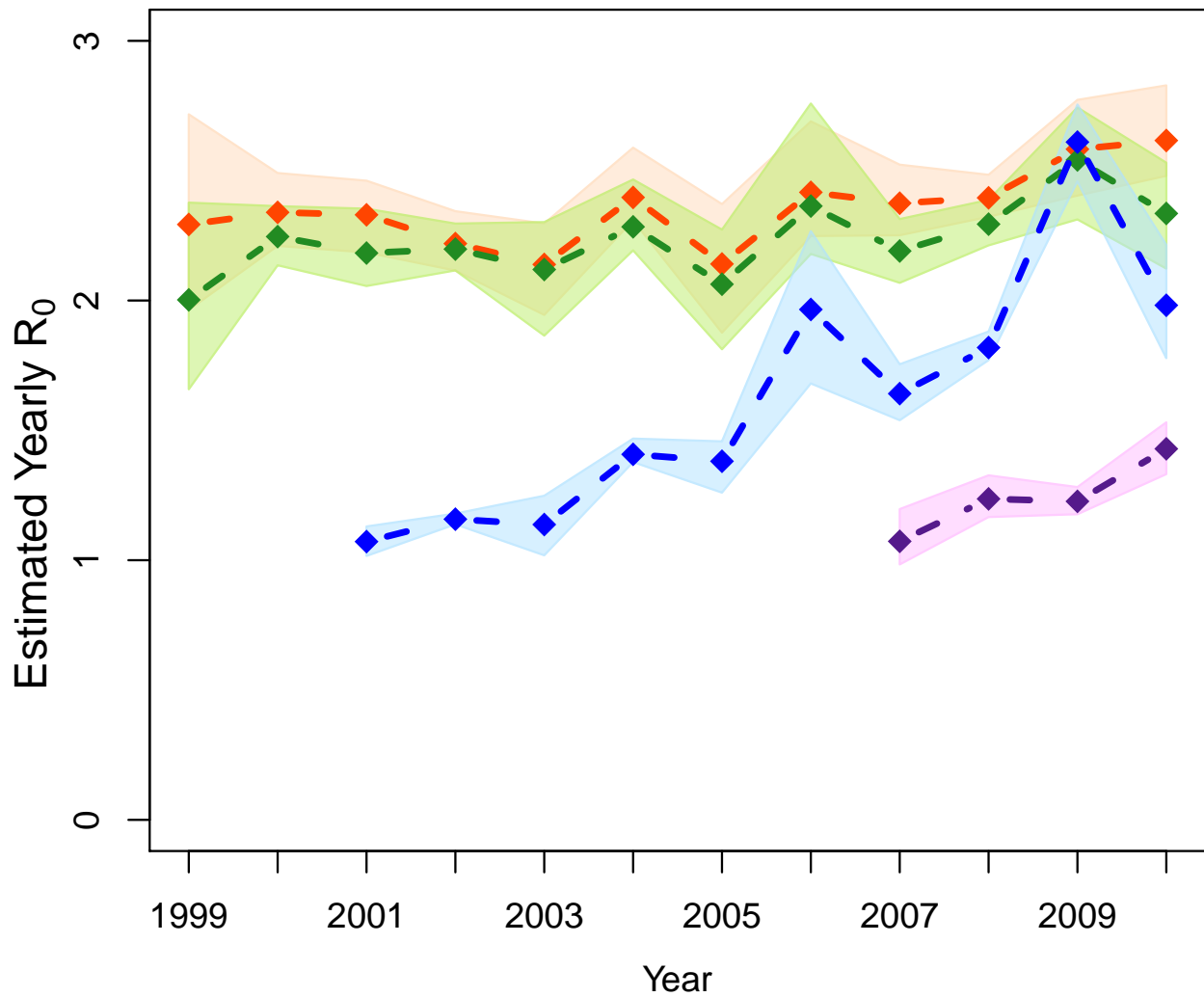
Fig. S32: **Yearly estimates of** $\mathcal{R}_0$. For each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), yearly estimates of $\mathcal{R}_0$ as well as the 90% BCI are plotted against time by computing the yearly mean of the daily estimates. The estimates for both DENV-3 and DENV-4 are truncated, excluding estimation before their respective introductions.
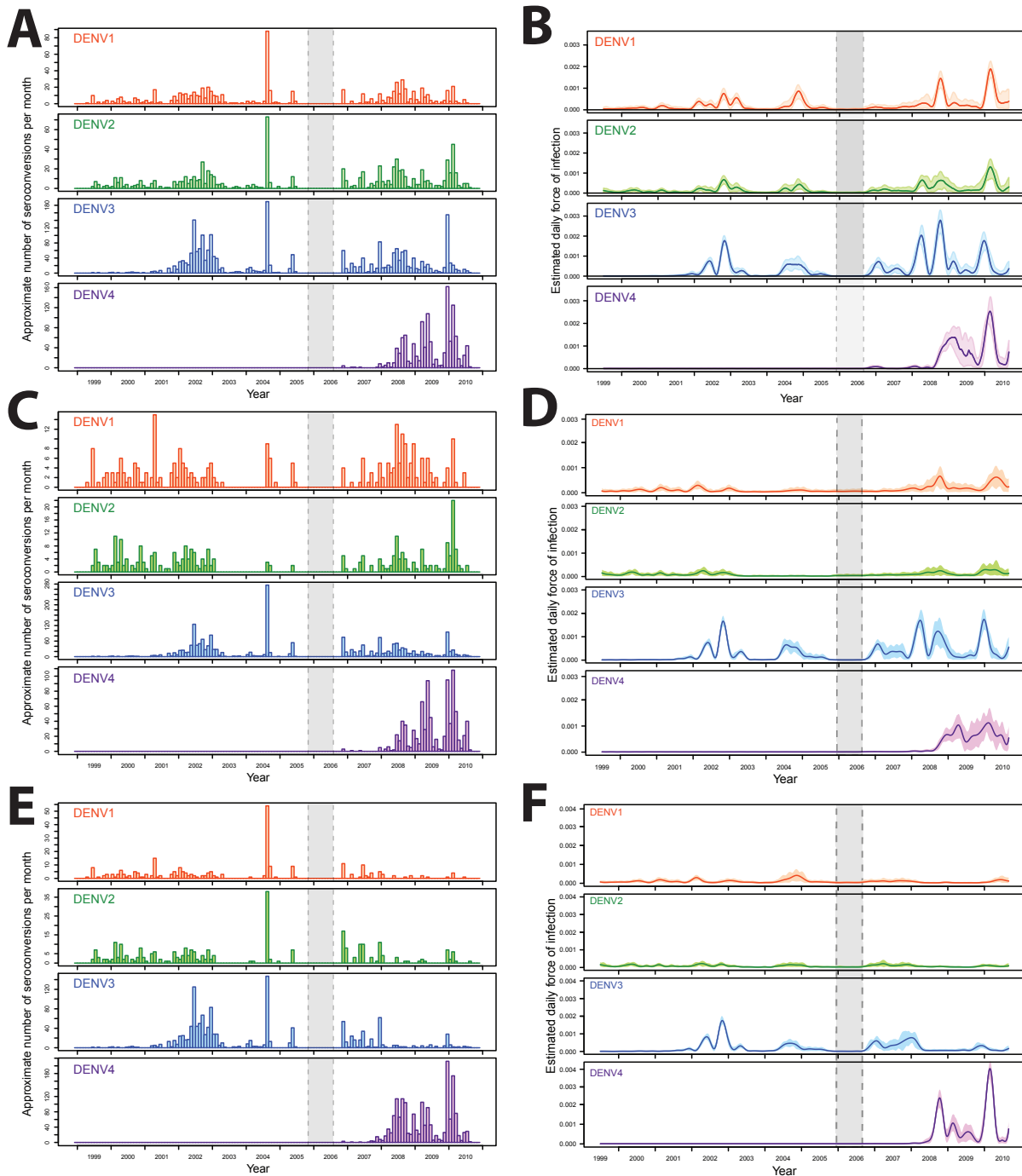
Fig. S33: **Results using three alternative data cleaning methods.** We repeat our analysis under different data cleaning approaches. Analogous to Figure 1 we plotted the number of interval censored infections by serotype against time, where the midpoint of the interval over which the infection was censored was used to time infections in (a), (c) and (e). Analogous to Figure 3, for each serotype (DENV-1 top panel (orange), DENV-2 second panel (green), DENV-3 third panel (blue) and DENV-4 bottom panel (purple)), daily estimates of the *FoI* as well as the 90% BCI are plotted against time in (b), (d) and (f). In (a) and (b) we used the least conservative cleaning method which allows for multiple seroconversions within any interval as well as automatically assuming any positive DENV-3 and DENV-4 test cannot be a false positive. In (c) and (d) we implement the restrictive assumption that any individual's first seroconversion from 2003 to 2008 was declared DENV-3 unless the individual already appeared to have seroconverted to DENV-3. In (e) and (f) we implement the restrictive assumption that any individuals' first seroconversion from 2008 to 2010 was declared DENV-4 unless the individual already appeared to have seroconverted to DENV-4.