Supplementary Materials

Distribution estimation of NAT-reactivity duration and the duration from initial NAT-reactivity to IgM-positivity, and subsequent updating of the annual incidence estimate of ZIKV infections for Puerto Rico, 2016

For human blood plasma, we estimate the distributions of the duration of NAT-reactivity of Zika virus and of the duration from NAT-detectability to IgM seroconversion, and we subsequently use the estimated means from these distributions to update the annual incidence estimate of ZIKV infections in Puerto Rico in 2016. A four-analysis sequence is used to accomplish this:

1. Estimate the doubling time of replication of Zika virus (ZIKV) in macaques from experimental data
2. Extrapolating the doubling time estimate found in 1 to humans, use observed (albeit censored) data from blood donors to model the distribution of the duration of NAT-reactivity; estimate the mean for the fitted model
3. Extrapolating the doubling time estimate found in 1 to humans, use observed (albeit censored) data from blood donors to model the distribution of the duration from NAT-detectability to IgM seroconversion (or IgM-positivity); estimate the mean for the fitted model.
4. Use the model-fitted distributions estimated in 2 and 3 to update the incidence estimates in Chevalier, *et al.*[1], 2017, where we augment the methods there by incorporating IgM-positivity status of donations to weight NAT-reactive, IgM-negative or IgM-positive donations differentially to reflect their relative likelihood of detection. Finally, NAT-reactive, IgM-positive donations are further weighted to reflect deferral of symptomatic (potential) donors.

Each of these analyses is undertaken in turn below, and we employ statistical resampling and bootstrapping in steps 2–4 to quantify uncertainty in estimated quantities and inferential statements.

Step 1: ZIKV Doubling Time in Macaques

Figure 1 shows the measured viral loads for 18 macaques infected with ZIKV on days 1-3 after inoculation on day 0, with the vertical axis being $\log(\text{Viral Load}/100)$ and the horizontal axis the day post-infection.

Because we are interested in estimating the doubling time, we need only the observations during the ramp-up phase before host immunity blunts viral replication, and this is evident in the data as the time before viral load decreases. These values are shown in Figure 2 as the filled circles, and these are the ones to which we restrict for the estimation of ZIKV doubling time. The following random-effects, linear model was used to fit a line to the ramp-up phase, transformed viral loads as a function of day, while accounting for potential within-macaque tendencies. For macaque $i = 1,2,\ldots,18$, with $j = 1,2,\ldots,n_i$ (here, $n_i = 3$ or 4) on days $d_{ij} = j - 1$, model the the observed pairs $\left(d_{ij}, ViralLoad_{ij}\right)$ as

$$\log(ViralLoad_{ij}/100) = (\delta + \Delta_i)\mathrm{d}_{ij} + \epsilon_{ij}$$

where $\Delta_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\Delta^2)$ are macaque-specific, random deviations from the fixed (average) slope δ, $\epsilon_{i0} \equiv 0$, and $\epsilon_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, for $j > 0$, are independent random errors, also independent of the Δs. Our interest is in the estimate of δ, which we use to estimate the ZIKV doubling time, which is $ZDT \equiv \log(2)/\delta$ days. Note that the scale factor 100 on the Viral Loads does not affect δ nor its estimate.

Fitting the mixed linear model given, the restricted maximum likelihood estimate (standard error) for δ is $\hat{\delta} = 3.11$ (0.21), and the fitted mean line with slope $\hat{\delta} = 3.11$ is shown as the dashed, gray line in the panels in Figure 1. The resulting estimate (asymptotic SE) for ZDT is $\hat{ZDT} = \log(2)/\hat{\delta} = 0.223$ (SE = $\log(2)SE_{\hat{\delta}}/\hat{\delta}^2 = 0.015$) days, or 5.35 (0.36) hours.

Step 2: Duration of NAT-reactivity

Weekly totals of Puerto Rico collections tested and the number of reactive donations.

Week	RR*(%)	No. RR	No. Tested
4/3/16	0.48%	5	1034
4/10/16	0.25%	3	1197
4/17/16	0.15%	2	1329
4/24/16	0.29%	4	1386
5/1/16	0.27%	3	1100
5/8/16	0.40%	5	1238
5/15/16	0.57%	8	1396
5/22/16	0.88%	12	1361
5/29/16	0.77%	9	1170
6/5/16	1.07%	17	1588
6/12/16	1.35%	21	1552
6/19/16	1.24%	17	1373
6/26/16	1.26%	18	1434
7/3/16	1.79%	19	1064
7/10/16	0.96%	13	1358
7/17/16	0.98%	13	1327
7/24/16	1.43%	15	1049
7/31/16	1.31%	16	1221
8/7/16	1.10%	16	1456
8/14/16	1.18%	17	1436
8/21/16	1.59%	21	1322
8/28/16	0.93%	17	1826
9/4/16	0.93%	13	1397
9/11/16	0.76%	13	1705
9/18/16	0.36%	4	1099
9/25/16	0.30%	5	1692
10/2/16	0.44%	7	1578
10/9/16	0.33%	5	1524
10/16/16	0.21%	3	1449
10/23/16	0.20%	3	1473
10/30/16	0.14%	2	1407
11/6/16	0.13%	2	1543
11/13/16	0.52%	8	1527
11/20/16	0.09%	1	1065
11/27/16	0.37%	6	1639
12/4/16	0.24%	4	1683
12/11/16	0.00%	0	1299
12/18/16	0.41%	4	969
12/25/16	0.00%	0	846
Total	**0.66%**	**351**	**53112**

*Repeat Reactive

Longitudinal results for 140 prospectively enrolled ZIKV NAT-reactive-at Index donors were obtained, including viral load, IgM and IgG results at Index donation, and NAT-reactivity at 2-3 follow-up times. Follow-up time 1

ranged roughly 5-77 days post-index (DPI), while follow-up time 2 ranged 15-72 DPI. A single donor provided a 3rd follow-up, at 42 DPI. Based on Index donation serology and viral load measurements, donors were classified into Stages, defined to reflect the relative timing of Index donation during the infection process, in which Stage I donors donated relatively soon after infection, while those in Stage V donated relatively late after infection.

To estimate the distribution of the duration of NAT-reactivity, we consider data for donors comprised of the time of the last NAT-reactive donation, L, and the time of the first NAT-non-reactive donation, U, as this interval provides for each donor bounds on the time of transition from NAT-reactivity to -nonreactivity. Data marking these intervals for the donors in Stages I-V are shown in Figure 2, in which + symbols mark endpoints of the intervals described. Intervals with both start- and end-times (so-called *interval censored*) are shown as segments connecting the + symbols, while intervals for donors who were not observed to revert to NAT-nonreactivity (whose observations are termed *right-censored*) are shown with + symbols at the last NAT-reactive observation and with arrowheads at the right edge of the plotting area. Donor Stage is indicated in Figure 2 by color, with Stage I donors' data grouped at the bottom of the graph, and Stages grouped in order up to those from Stage V displayed at the top. Note that oftentimes the last positive donation was at index, so that $L = 0$, while for right-censored observations, $U = \infty$, by convention.

A standard statistical modeling approach used for time-to-event data is survival analysis, and estimation methods accommodate censoring. We use parametric survival models described below. If infection had occurred coincident with the day of Index donation, modeling could proceed with the data available, however, donors' presentation for donation only rarely would coincide with the date of infection. Since all donors in this study were NAT-reactive on Index donation, having been infected at some unknown, previous time, time of infection is *left-censored*. To model the full duration of NAT-reactivity, not truncated by this left-censoring, we use the observed viral load measurements for the donors obtained from the Index donations to impute time of initial NAT-reactivity, which is expected to be within hours of infection, using the ZIKV doubling time estimated for macaques in Step 1. To do this, recall that the $ZDT = \log(2)/\delta$ means that the viral load at time t after infection, $\Lambda(t)$, is

$$\Lambda(t) = \Lambda(0)\, 2^{t/ZDT}$$

for initial (minimum NAT-detectable) viral load $\Lambda(0)$. From this expression, for observed viral load $\Lambda_i(t_o)$ for donor i at unknown time t_{oi} after infection, we have

$$\begin{aligned} t_{oi} &= ZDT[\log_2\Lambda_i(t_{oi}) - \log_2\Lambda_i(0)] \\ &= \frac{\log(2)}{\delta}[\log_2\Lambda_i(t_{oi}) - \log_2\Lambda_i(0)] \end{aligned}$$

We therefore may, using this equation with estimated values for the parameters, impute times of infection from observed viral loads, assuming minimum detectable NAT concentration $\Lambda_i(0)$. A common estimate $\widehat{\Lambda}(0)$ = 7.85 (95% CI 6.21, 9.92) IU/mL used for each donor was obtained as the geometric mean of the individual Roche assay [3.9 (95% CI 2.8, 5.3) IU/mL] and Vitalant confirmatory PCR assay [15.8 (95% CI 11.2, 22.2) IU/mL] 50% LOD estimates. We incorporated uncertainty in the estimate $\widehat{\Lambda}(0)$ by resampling for each imputation iteration below from a normal distribution with mean 2.06 and standard deviation 0.12, then exponentiating.

Because we use the ZDT estimate from the macaque study, which was derived only from observations during the ZIKV ramp-up phase (that is, before antibody begins to retard ZIKV replication), we estimate the NAT-detectability duration using only those donors in Stages I and II, shown in Figure 2, who by definition of the stages satisfy this same constraint. For each donor, $i = 1,2, ..., 90$, we impute $s = 1,2, ..., 10{,}000$ values $t_{oi}^{(s)}$ of t_{oi}, and these values are used as imputed observation intervals $I_i^{(s)} \equiv \left(t_{oi}^{(s)} + L_i, t_{oi}^{(s)} + U_i\right)$ bounding the times of reversion from NAT-reactivity to NAT-nonreactivity. For comparison, Figure 3 shows in the left panel the original intervals for donors in Stages I and II (as from Figure 2), and in the right panel are shown these same data with estimated mean imputed time from infection detectable by NAT, $\overline{t_{oi}^{(\cdot)}}$, to date of Index donation incorporated, displayed in the middle panel of Figure 3.

We use standard survival analysis methods for interval censored data to fit both Weibull and log-normal distributions to the imputed NAT-detectability intervals $\left\{I_i^{(s)}\right\}_{i=1}^{90}$, for each $s = 1,2, ..., 10{,}000$, to account for the

uncertainty in the imputations of the durations from infection to Index donation. For each simulation *s*, model parameters were estimated using maximum likelihood, and these estimates were used to compute estimates of the underlying distribution means. Further, for each simulation *s*, Akaike's Information Criterion (AIC) was computed and used to compare model fits; models with lower AIC values are preferred.

Of the 10,000 simulations, the AIC was lower for the log-normal distribution than for the Weibull for 99.13% of the simulations, and the average difference in AIC was -12.19. We therefore continue only with the log-normal distribution. Estimates of means (over the simulations) of the log-normal parameters were log-scale mean = 2.34 and log-scale standard deviation = 0.5. The median duration was estimated as 10.33 (95% CI 9.16-11.84) days and the 5th and 95th percentiles of the durations were 4.55 and 23.47 days, respectively. A histogram of the simulated values of the means from the fits is shown in Figure 4. The 95% confidence interval for the mean NAT-detectability duration 11.70 days is computed from the 2.5th and 97.5th quantiles of these simulated values as (10.06 - 14.36) days.

Step 3: Duration from initial NAT-detectability to IgM-positivity

With imputations for commencement of NAT-detectability as a surrogate for timing of infection in-hand, we may employ these to estimate the duration from initial NAT-reactivity to IgM-positivity using the same statistical approach. In addition to NAT, each donation was tested for IgM antibodies to ZIKV, including at Index donation. As above, we restrict this analysis to donors in Stages I and II to be able to use the viral load data to impute the initial time of NAT-reactivity. IgM is negative at Index by definition for these stages. Figure 5 shows for these IgM data the original intervals in the left panel, the estimated average time from NAT-detectability to donation by observed viral load in the middle panel, and the updated IgM seroconversion intervals in the right panel.

As with the NAT-reactivity duration analysis, a majority (54.94%) of the simulations, the AIC was lower for the log-normal distribution than for the Weibull, and the average difference in AIC was -4.85. We therefore continue only with the log-normal distribution. Estimates of means (over the simulations) of the log-normal parameters were log-scale mean = 2 and log-scale standard deviation = 0.19. The estimated median duration from initial NAT-reactivity to IgM-positivity was 7.42 (95% CI 6.59 – 8.29) days, and the 5th and 95th percentiles of this duration were 5.39 and 10.21 days, respectively. A histogram of the simulated values of the means from the fits is shown in Figure 6. The estimate of the mean duration from initial NAT-reactivity to IgM-positivity is thus 7.56 days, and the computed 95% confidence interval is obtained as the 2.5th and 97.5th percentiles of these simulated values, giving (6.74 - 8.39) days.

The results of this and the previous section are succinctly summarized in Figure 7 (Figure 3 of the manuscript), where the distribution estimates of both the duration of NAT-reactivity and of duration from initial NAT-reactivity to IgM-positivity are displayed. These are indicated as probability of detection curves by time (with 95% confidence bands), with NAT (blue) naturally decreasing over time, and IgM (red) increasing over time. The top panel of the figure reproduces these distributional estimates as densities, to facilitate interpretation of the distributional shapes for these characteristics. Marked as vertical line segments in the densities are the medians (solid) and means (dashed) for the two distributions.

Step 4: Puerto Rico ZIKV Incidence, 2016

In Chevalier, *et al.*[1], estimates of ZIKV infection incidence were computed for the time period April 3, 2016, through August 12, 2016, using observed weekly blood donor positivity rates, stratified by health region, age group, and sex. The breakout of the demographics of the screened population is presented in the following table.

Age Group	Female		Male		Total	
	N	%	N	%	N	%
16-29	4,713	30.4	5,831	20.8	10,544	24.3
30-44	4,942	31.9	9,088	32.5	14,030	32.3
45-59	4,257	27.5	9,103	32.5	13,360	30.7
60-74	1,516	9.8	3,674	13.1	5,190	11.9
75+	68	0.4	272	1.0	340	0.8
Total	15,496	100	27,968	100	43,464	100

Estimation relied on specification of parameters characterizing the mean duration of NAT-reactivity, asymptomatic infection proportion, and mean incubation period. As noted there, reported incidence estimates relied heavily on the specification of the mean duration of NAT-reactivity assumed, and though this was set at 9.9 (95% CI 6.8–21.6) days, estimates were reported for a range of this parameter from 7 to 21 days (Figure 3 in Chevalier, *et al.*[1]).

We employ the general estimation approach detailed in the supplementary material for Chevalier, *et al.*[1], but our analysis here updates the reported, final incidence estimate in Chevalier, *et al.*[1], by:

(a) using data now available for the full 2016 year,
(b) using our updated estimate of the mean duration of NAT-reactivity reported above in Step 2,
(c) using our estimate of the mean duration from NAT-detectability to IgM seroconversion in Step 3, and finally
(d) by using the available IgM-positivity status of the NAT-reactive donors.

In Chevalier, *et al.*[1], the expected (mean) duration of NAT-detectability for potential donors was estimated, which in previous publications (e.g., Busch, *et al.*[2]) did not account for deferral of symptomatically infected persons when estimating this quantity. Further, in neither Chevalier, *et al.*[1], nor in Busch, *et al.*[2], were estimates of the numbers of infected individuals in the population inflated to account for ZIKV-infected, symptomatic individuals in the population, who are not directly reflected among donors because donors are required to be asymptomatic. The availability of estimates of the mean durations from initial NAT-detectability to IgM seroconversion (dIgM), and from initial NAT-detectability to the loss of NAT-detectability (dNAT) can be used as follows to refine the estimation approach in Chevalier, *et al.*[1], in which donor IgM seroconversion status was unknown. We assume that for ZIKV-infected individuals who develop symptoms that the onset of symptoms effectively coincides with IgM seroconversion. For estimating incidence, we separated the NAT-reactive donors into pre- and post-IgM seroconversion groups and estimated the numbers of ZIKV infected individuals in the population for these categories using different NAT-detectability weights: the reciprocal of dIgM for the former; and the reciprocal of the mean duration from IgM-seroconversion to loss off NAT-detectability for the latter, where this was estimated as the difference dNAT–dIgM. Here we assumed that all individuals who were NAT-reactive and IgM-negative represent all population infections in the pre-IgM seroconversion phase, and these are assumed to be asymptomatic at the conceptual time of "observation" reflected by this donor group in the data. For the second group, we estimated directly all individuals who were NAT-reactive and IgM-positive and asymptomatic at the conceptual time of "observation" reflected by this donor group in the data; we then inflated this group's estimate by weighting by the reciprocal of the asymptomatic proportion to count the conceptual symptomatic individuals in the population who would be NAT-reactive and IgM-positive, but who are necessarily not represented in the donor data because donors must be asymptomatic to donate.

Figure 8 shows the numbers and proportions of NAT-reactive donations by week. There were 45,417 donations from the two collaborating blood collection agencies, of which 307 (0.7%) were ZIKV (cobas) NAT-reactive. Of these 307 NAT-reactive donors, 107 (34.9%) were IgM-positive.

Estimation of the total number of ZIKV-infected individuals in Puerto Rico in 2016 follows the methodology of Chevalier, *et al.*[1], with the following changes. First, we added the use of bootstrapping to resample the donations, which reflects a view of the present modeling effort as one modeling the underlying population transmission process (as with Biggerstaff and Petersen for WNV[3]), thereby reflecting essentially the uncertainty in random donor infection and donation/participation. Second, as noted, we use the updated estimates for NAT-reactivity duration computed above. To account for both the bootstrap uncertainty, the uncertainty in estimation of the ZDT ($\hat{\delta}$), the

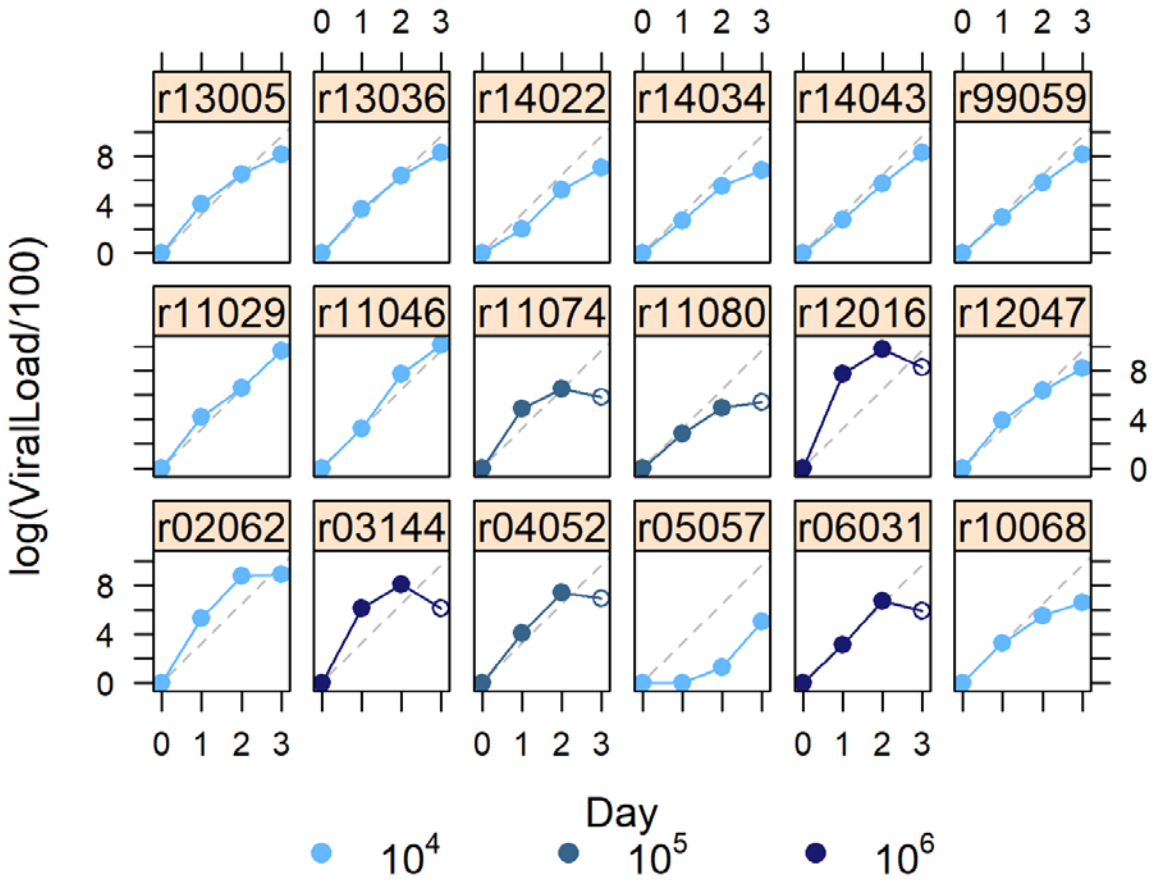
uncertainty in estimating the NAT-reactivity and duration, and the uncertainty in estimating the duration from NAT-detectability to IgM seroconversion, for each of the 9,946 bootstrap samples of the donor results we used one of the 9,946 sampled values of the log-normal means shown in Figures 4 and 6 to compute an estimate of the number of ZIKV infections. We next averaged the resulting estimated number of ZIKV infections to produce the final estimate, while using the 2.5th and 97.5th percentiles of these values as 95% confidence limits for the population number of ZIKV infections. Finally, we divided the estimate (95% CI) by the population size of Puerto Rico to give the estimate of the incident number of ZIKV infections for all of 2016. Other parameter specifications for the resampling estimation remain the same as in Chevalier, *et al.*{Chevalier, 2017 #84}, namely that the mean incubation period was assumed to be normal-distributed with mean 6.20 and standard deviation 0.48 days, and that the mean asymptomatic rate was assumed to be beta distributed with mean 0.79 and standard deviation 0.043.

A histogram of the resample values for the number of ZIKV infections is shown in Figure 9, where the estimate of 768,101 (95% CI 659,298 – 878,702) ZIKV infections is shown. With population size 3,638,773, this gives an estimated incidence for 2016 of 21.1% (95% CI 18.1% – 24.1%).

References

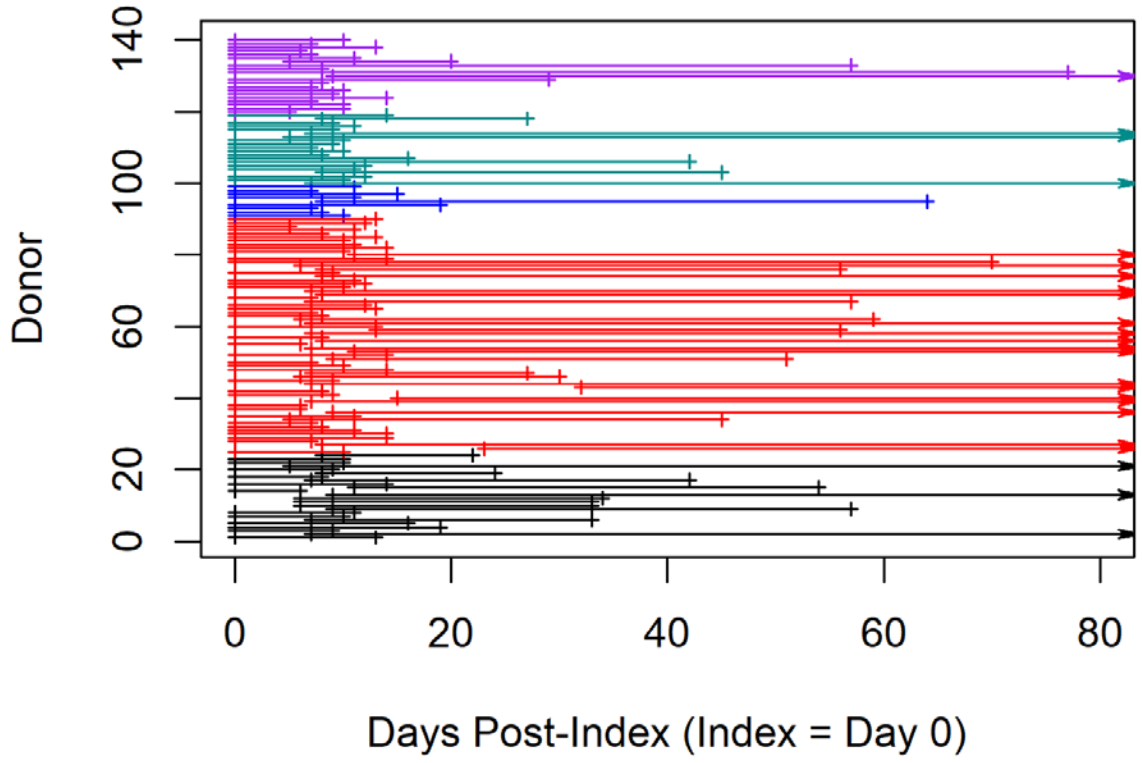
1. Chevalier MS, Biggerstaff BJ, Basavaraju SV, et al. Use of Blood Donor Screening Data to Estimate Zika Virus Incidence, Puerto Rico, April-August 2016. *Emerg Infect Dis* 2017; **23**(5): 790-5.
2. Busch MP, Kleinman SH, Tobler LH, et al. Virus and antibody dynamics in acute west nile virus infection. *J Infect Dis* 2008; **198**(7): 984-93.
3. Biggerstaff BJ, Petersen LR. Estimated risk of West Nile virus transmission through blood transfusion during an epidemic in Queens, New York City. *Transfusion* 2002; **42**(8): 1019-26.

Figure 1: log(Viral Load / 100) by Day for 18 experimentally infected macaques.



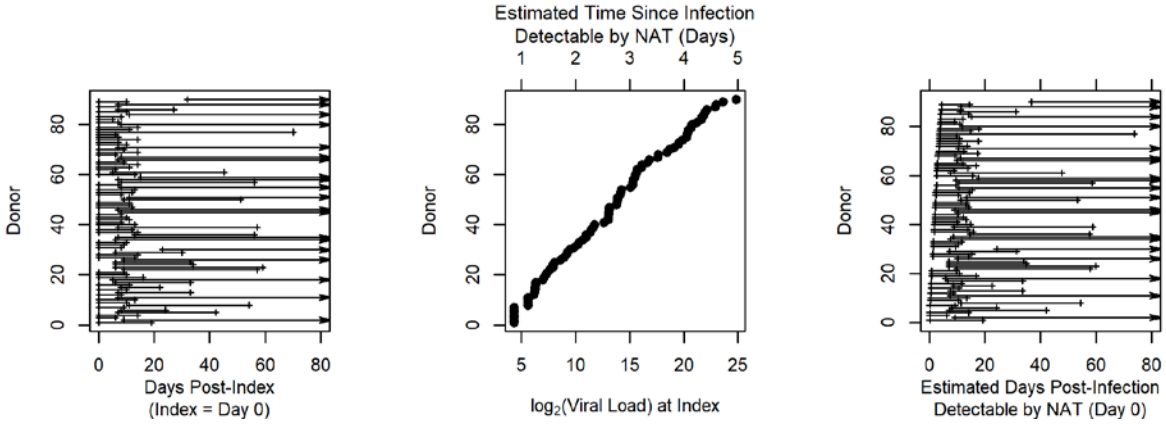
ID numbers shown in panel banners. Dose (virus particles per mL) is indicated by colors, identified in the legend. Only the filled circles are used in the estimation of the doubling time. The fitted mean line is shown as the dashed, gray line.

Figure 2: Observed intervals capturing time of NAT reactivity-to-nonreactivity events for analysis



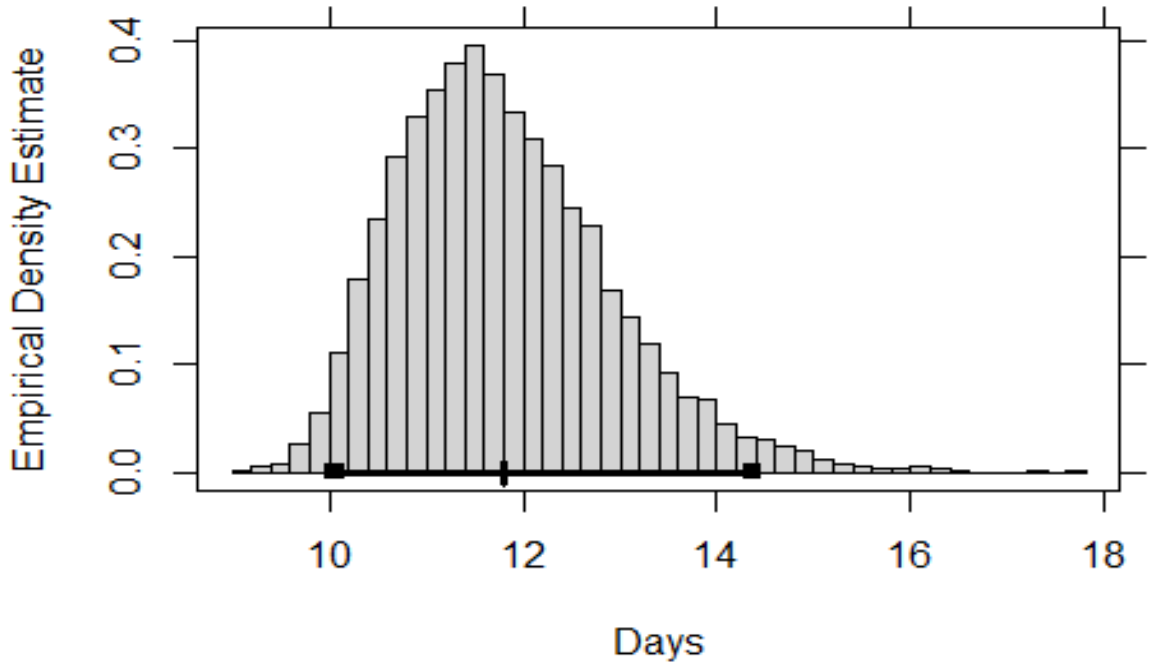
Observed intervals area colored by Stage (I = black, II = red, III = blue, IV = cyan, V = purple). Arrowheads at the right edge of the plotting area indicate right-censored data, for which reversion to NAT-nonreactivity was not observed.

Figure 3: Intervals capturing NAT reactivity-to-nonreactivity events for analysis and observed viral loads.



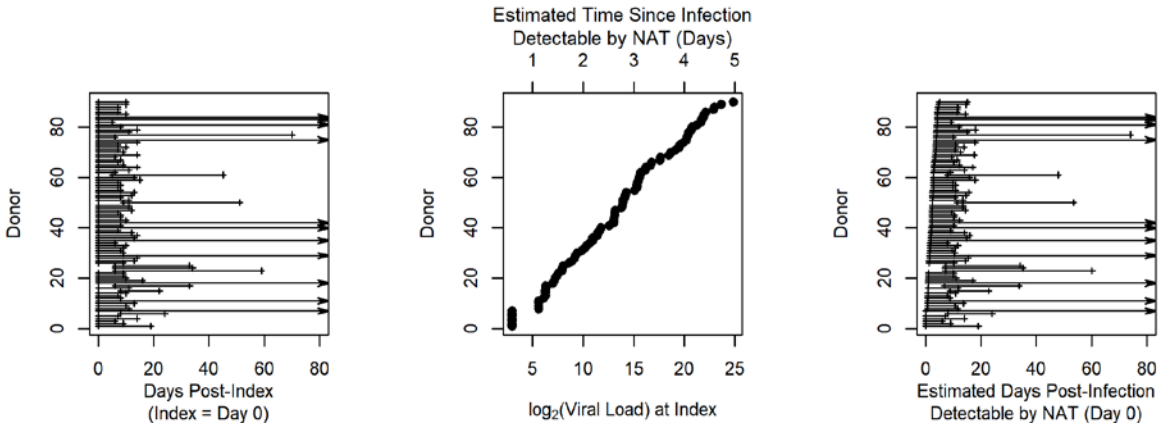
The left panel contains observed intervals, using Days Post-Index; the middle panel shows the log2-observed viral loads (bottom axis) and the corresponding estimated time in days since infection detectable by NAT (top); the right panel are the observed intervals adjusted to (Mean) Estimated Days Post-Infection Detectable by NAT.

Figure 4: Histogram of the simulated values of the mean of the log-normal distribution fits to the NAT-reactivity duration data, incorporating imputation of duration from initial NAT-reactivity to Index donation.



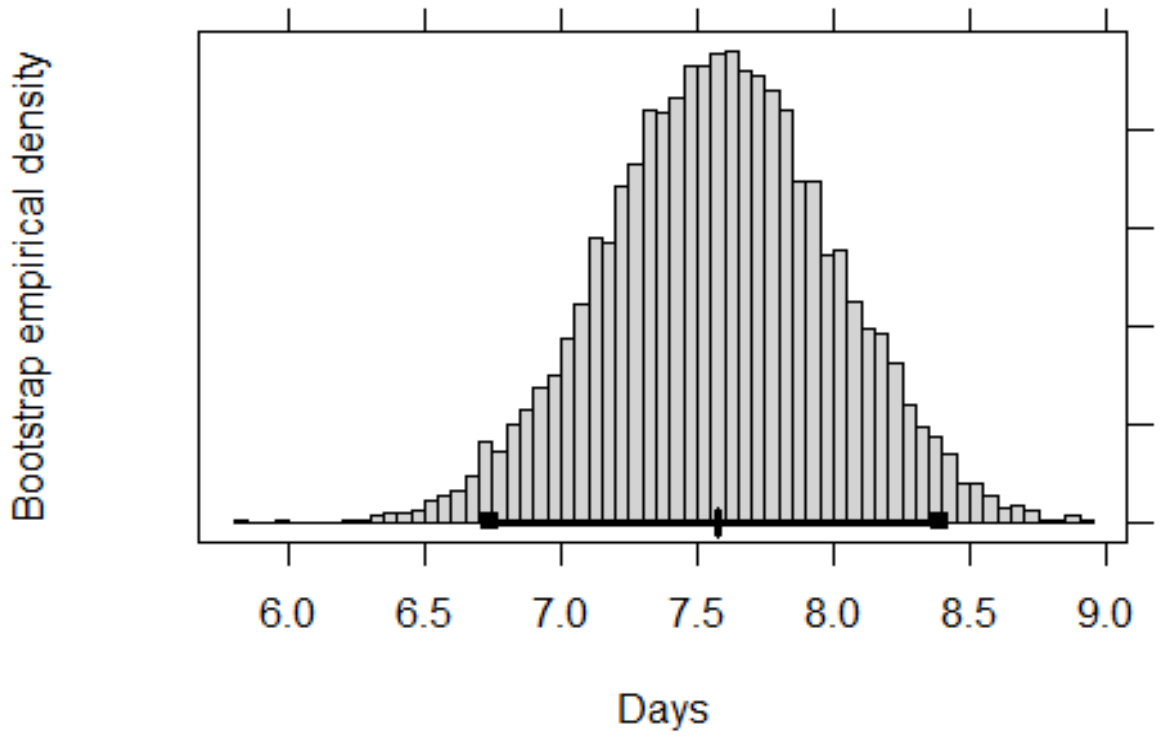
The mean and 95% confidence interval for the mean are indicated with the segment along the bottom of the histogram.

Figure 5: Intervals capturing IgM negativity-to-positivity events for analysis and observed viral loads.



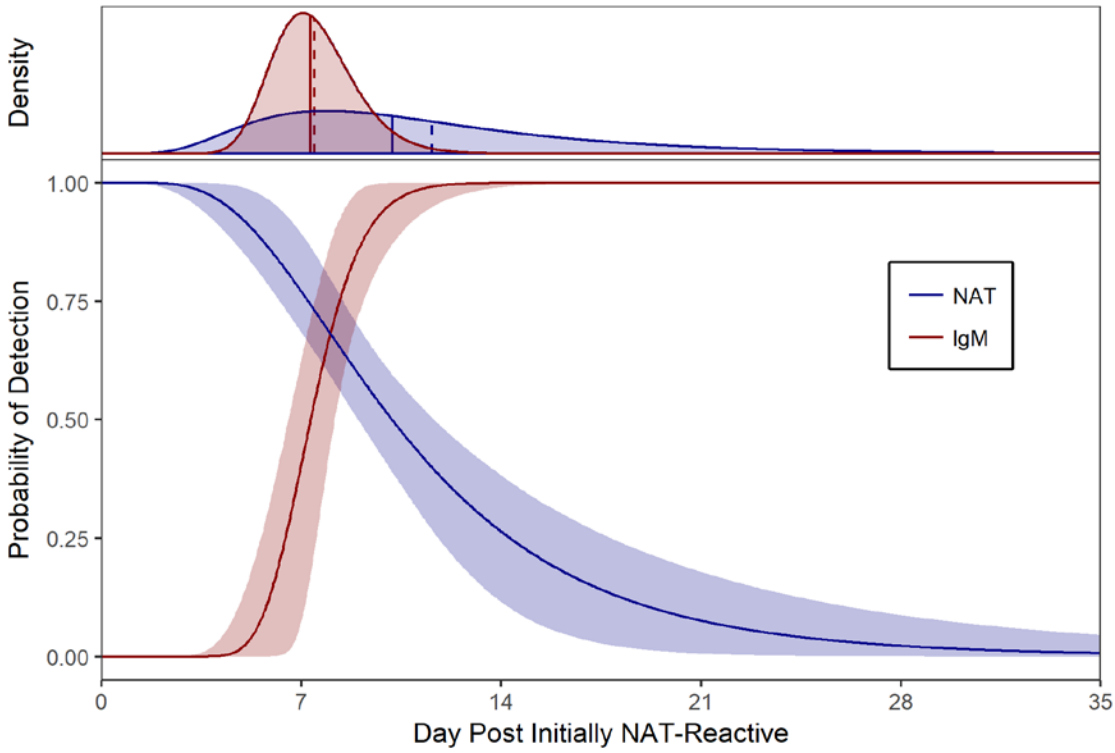
The left panel contains the observed intervals, using Days Post-Index; as above, the middle panel shows the log2-observed viral loads (left axis) and estimated time in days since infection detectable by NAT (right axis); the right panel contains the observed intervals adjusted to (Mean) Estimated Days Post-Infection Detectable by NAT to IgM-positivity.

Figure 6: Histogram of the simulated values of the mean of the log-normal distribution fits to the duration from initial NAT-reactivity to IgM-positivity data, incorporating imputation of duration from NAT-reactive infection to Index donation.



The mean and 95% confidence interval for the mean are indicated with the segment along the bottom of the histogram.

Figure 7: In the bottom panel are fitted probabilities of detection curves over time for the duration of NAT-reactivity (blue) and of the duration from initial NAT-reactivity to IgM-positivity (red), with 95% confidence bands shown shaded in corresponding colors. The top panel of the figure reproduces these distributional estimates as densities, to facilitate interpretation of the distributional shapes for these characteristics.



Marked as vertical line segments in the densities are the medians (solid) and means (dashed) for the two distributions.

Figure 8: Percent and number of ZIKV NAT-reactive donations by week, Puerto Rico, 2016

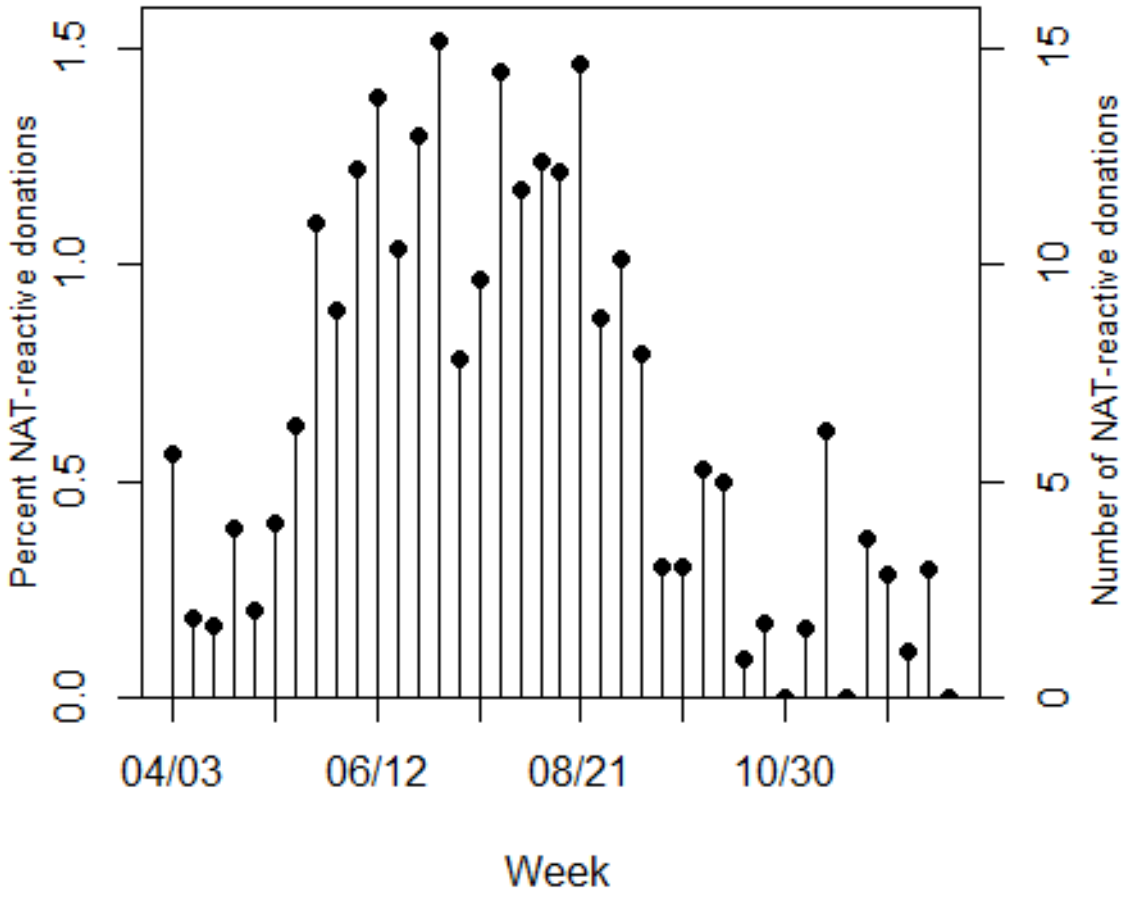
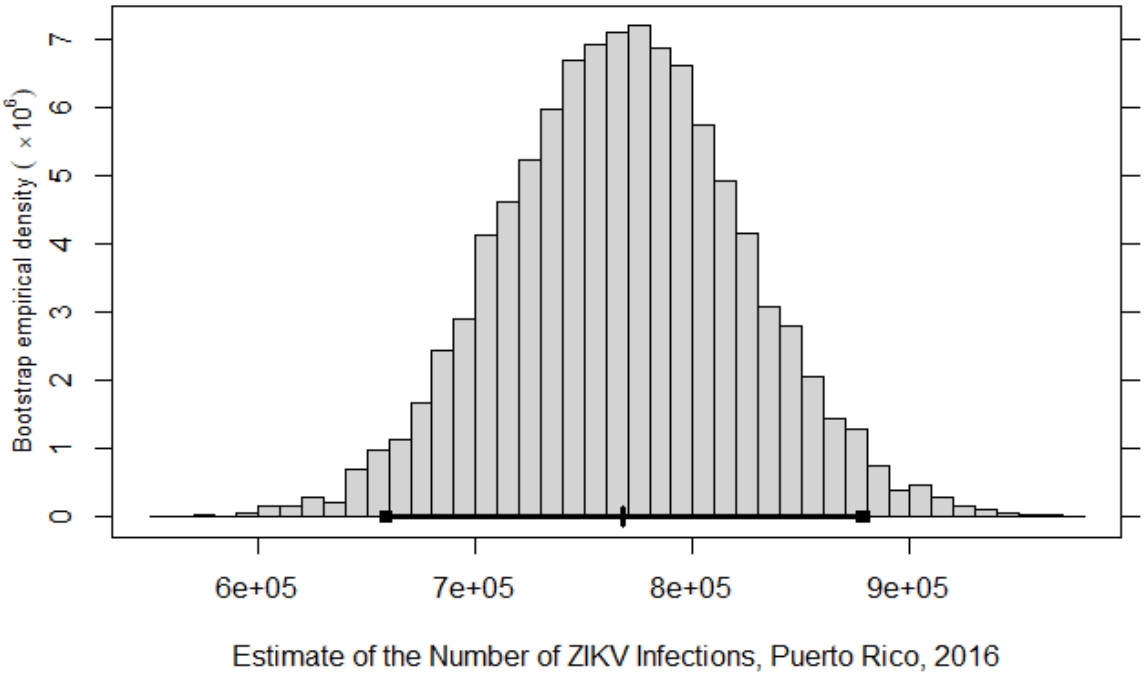


Figure 9: Histogram of the resample estimate values for the number of ZIKV infections, Puerto Rico, 2016.



The interval in black represents final estimate (mean) and 95% confidence interval for the number of ZIKV infections.