

## **Estimating incidence of infection from diverse data sources: Zika virus in Puerto Rico, 2016**

Talia M. Quandelacy<sup>1</sup>, Jessica M. Healy<sup>2</sup>, Bradford Greening<sup>2</sup>, Dania M. Rodriguez<sup>1</sup>, Koo-Whang

Chung<sup>2</sup>, Matthew J. Kuehnert<sup>2</sup>, Brad J. Biggerstaff<sup>3</sup>, Emilio Dirlikov<sup>2</sup>, Luis Mier-y-Teran-Romero<sup>1</sup>, Tyler

M. Sharp<sup>1,4</sup>, Stephen Waterman<sup>1,4</sup>, Michael A. Johansson<sup>1,5</sup>

### **Affiliations**

1. Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San Juan, PR, USA

2. Centers for Disease Control and Prevention, Atlanta, GA, USA

3. Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, CO, USA

4. United States Public Health Service, Silver Springs, MD, USA

5. Harvard TH Chan School of Public Health, Boston, MA, USA

### **Corresponding Author**

Talia M. Quandelacy (tquandelacy@cdc.gov)

## Abstract

Emerging epidemics are challenging to track. Only a subset of cases is recognized and reported, as seen with the Zika virus (ZIKV) epidemic where large proportions of infection were asymptomatic. However, multiple imperfect indicators of infection provide an opportunity to estimate the underlying incidence of infection. We developed a modeling approach that integrates a generic Time-series Susceptible-Infected-Recovered epidemic model with assumptions about reporting biases in a Bayesian framework and applied it to the 2016 Zika epidemic in Puerto Rico using three indicators: suspected arboviral cases, suspected Zika-associated Guillain-Barré Syndrome cases, and blood bank data. Using this combination of surveillance data, we estimated the peak of the epidemic occurred during the week of August 15, 2016 (the 33<sup>rd</sup> week of year), and 120 to 140 (50% credible interval [CrI], 95% CrI: 97 to 170) weekly infections per 10,000 population occurred at the peak. By the end of 2016, we estimated that approximately 890,000 (95% CrI: 660,000 to 1,100,000) individuals were infected in 2016 (26%, 95% CrI: 19% to 33%, of the population infected). Utilizing multiple indicators offers the opportunity for real-time and retrospective situational awareness to support epidemic preparedness and response.

## Introduction

The emergence and rapid spread of Zika virus (ZIKV), an arbovirus transmitted by *Aedes* species mosquitoes, in the Americas (1) resulted in large-scale epidemics throughout the tropical areas of the region. The first confirmed locally acquired ZIKV case in Puerto Rico was reported on December 31, 2015 (2), followed by more than 36,000 confirmed cases in 2016 (3). While confirmed cases provided an indicator of transmission intensity, reported cases represented a small proportion of actual infections (4) in part because many ZIKV infections are asymptomatic or mild, and are not captured by surveillance systems (5–7). Furthermore, distinguishing symptomatic (i.e. disease) cases of ZIKV infections from other arboviral infections (e.g., dengue, chikungunya) was difficult due to their similar symptoms (e.g. fever, rash), and serological cross-reactivity with dengue viruses (DENV). Despite these challenges, estimating the underlying ZIKV infection incidence was critical to assess useful metrics (e.g., transmission intensity, the number of people previously infected, and the number still at risk) that

informed prevention and response measures, and preparation for severe outcomes like Guillain-Barré Syndrome (GBS) (8) and congenital Zika syndrome (9).

ZIKV serosurveys, like those conducted in Yap and French Polynesia (5,10), provided estimates of cumulative incidence, but are logistically difficult, require substantial time and resources, and present diagnostic challenges due to varying duration of infection markers (RNA and different types of antibodies) and cross-reactivity (11). Therefore, it is important to find alternative methods to estimate incidence of infection, including statistical techniques that can be easily applied to surveillance data.

Although many infections are undetected during outbreaks, data exist for the set of infections captured through surveillance systems. Bayesian statistical methods explicitly consider both variability in observations (data) and uncertainty in model parameters (e.g., probability of observation), and are well-suited to address challenges like estimating quantities that are not directly observed. During the emergence of ZIKV in Martinique, *Andronico et al.* (12) developed a Bayesian model to explicitly incorporate a classic epidemiological compartmental model with surveillance data from Martinique using prior information on ZIKV transmission, reporting rates, and GBS risk from French Polynesia. We employed a similar approach in Puerto Rico incorporating multiple surveillance indicators and prior information on the probability of observing infections.

We considered surveillance data on suspected arbovirus cases, suspected Zika-associated GBS cases, and infections identified through a subset of blood banks as indicators of infection. Suspected arbovirus cases identified through passive surveillance reflect symptomatic care-seeking individuals with symptoms indicative of ZIKV, dengue virus, or chikungunya virus infection. Suspected Zika-associated GBS cases represented a more severe and easily recognized manifestation, though GBS can also result from other causes. Blood donor data provided information on asymptomatic and pre-symptomatic infections identified through blood screening. To capture underlying infection dynamics, we used a generic Time-series Susceptible-Infected-Recovered epidemic model within a Bayesian framework to relate infections to data by utilizing evidence-based assumptions on detection probabilities for each

indicator. In this framework, we estimated the weekly incidence on ZIKV infection and the cumulative number of infections in Puerto Rico in 2016.

## Methods

### *Data*

We collected data on suspected arboviral disease cases, suspected GBS cases, and infections detected among blood donors reported in Puerto Rico during January 1, 2016–December 31, 2016. A suspected arboviral disease case was defined as any patient with clinically-suspected illness resulting from an arbovirus infection and reported through the Passive Arboviral Diseases Surveillance System (PADSS) in Puerto Rico (3,13). Suspected GBS cases were patients experiencing onset of neurological symptomatic characteristic of GBS (e.g., bilateral flaccid limb weakness (14)) reported through the GBS Passive Surveillance System—a surveillance system capturing GBS cases along with patients experiencing other neurological symptoms (e.g., encephalitis (14)) and operated by the Puerto Rico Department of Public Health (PRDH) with support from the Centers for Disease Control and Prevention (CDC) (15). Not all suspected GBS cases were confirmed to be either ZIKV infections or GBS cases, rather they represented real-time reports of possible GBS cases. We used an indicator for asymptomatic and pre-symptomatic infections using blood bank data beginning in April 2016 when all blood donations were tested for ZIKV RNA (16). Two blood collection agencies provided the numbers of total donations and the number testing positive for ZIKV (16), the population of donors was not representative, being predominantly male and not including anyone under age 16. For the population of Puerto Rico, we assumed the population was approximately 3.4 million people based on 2016 census estimates (17).

### *Epidemic model*

We developed a generalized Bayesian discrete Time-series Susceptible-Infected-Removed model to fit surveillance data over the 52 weeks of 2016 to estimate weekly ZIKV infection incidence in Puerto Rico. We used an underlying SIR epidemic model, where a proportion of the population was infected each week,  $z_t$ , and was defined as the product of the proportion infected in the previous week ( $z_{t-1}$ ), the

proportion of population that was susceptible in the previous week ( $s_{t-1}$ ), and a time-varying transmission rate,  $\beta_t$ :

$$z_t = \beta_t * z_{t-1} * s_{t-1}$$

$$\beta_t \sim \text{Normal}_{(0,\infty)}(\beta_0, \sigma)$$

where we use the notation  $\text{Distribution}_{(a,b)}(\theta)$  to indicate that we use the stated Distribution with relevant parameter(s)  $\theta$  but restricted to support (a,b), and so scaled to yield a valid distribution. The time-varying transmission rate ( $\beta_t$ ) was assumed to have a constant mean reflecting no substantial control, with random variability between weeks, and was constrained to be greater than or equal to zero using a half-normal distribution. We assumed a prior for the transmission rate ( $\beta_0, \beta_0 \sim \text{Normal}_{(0,\infty)}(2,1)$ ) reflecting an expected weekly reproductive number on the order of 1 to 5 (18). The prior for the standard deviation had a similar scale ( $\sigma \sim \text{Normal}_{(0,\infty)}(0,1)$ ). Though the average time between successive generations of arboviral infections, i.e., generation times, are typically several weeks (19,20), we implemented this model with a weekly time step, intended to reflect a generic representation of a weekly transmission process in which  $\beta_t$  cannot be directly interpreted as  $R_0$ , the basic reproductive number. We did not explicitly aim to model the initial phases of the epidemic, and therefore only estimated an initial proportion of the population infected in the first week of 2016, using a restricted normal prior to indicate a small prevalence of infection that week ( $z_0 \sim \text{Normal}_{(0,1)}(0,0.001)$ ). All other individuals were assumed susceptible to infection, as evidence suggests only very limited transmission prior to 2016 (2). Our model assumed a closed population, meaning that susceptible and infection population estimates depended only on population-level risk, and that there were no births, deaths, or migration.

### *Reporting models*

For each surveillance indicator, we estimated the probability of observing an infection as a function of infection risk ( $z_t$ ) and the observation process for each data type within the epidemic model.

The combined model estimated incident infections from the three individual indicators within one epidemic model.

### *Suspected arboviral cases*

Given that laboratory testing identified very few dengue and chikungunya cases (3), we assumed most suspected arboviral cases were suspected Zika cases. We estimated the expected number of reported suspect arboviral cases as the product of population size ( $N$ ), ZIKV infection prevalence ( $z_t$ ), and the probability of reporting a clinical suspect Zika case per infection ( $p_{S|Z}$ ) and fit the case data using the negative binomial distribution formulated as a mean and dispersion:

$$S_t \sim \text{NegBin}(N * z_t * p_{S|Z}, \varphi_S).$$

For  $p_{S|Z}$ , we used a beta-distributed prior to approximate a mean of 0.11 and 95% credible interval [CrI] of 0.01-0.24 based on Mier-y-Teran et al. (21) ( $p_{S|Z} \sim \text{Beta}(3.3, 27)$ ) (Table 1). The dispersion,  $\varphi_S$ , was assigned a prior distribution with high expected overdispersion,  $\varphi_S \sim \text{Normal}_{(0,\infty)}(0, 1000)$ .

### *Suspected ZIKV-associated GBS cases*

We assumed the number of observed suspected GBS cases ( $GBS_t$ ) came from a binomial distribution:

$$GBS_t \sim \text{Binomial}(p_{G0} + p_{G|Z} - p_{G0} * p_{Z(t)}, N),$$

where  $p_{G0}$  is the weekly risk of GBS due to other causes and  $p_{G|Z}$  is the probability of suspect GBS given ZIKV infection and  $p_{G0} + (p_{G|Z} - p_{G0} * p_{Z(t)})$  represents the probability of an individual in population  $N$  being reported as a suspected GBS case. We assumed the global baseline GBS risk was 0.8-1.9 GBS cases per 100,000 per year (22) for Puerto Rico, and the weekly risk was approximated using a beta prior distribution:  $p_{G0} \sim \text{Beta}(23, 8.9 \times 10^7)$ . The prior for the probability of suspected GBS given ZIKV infection  $p_{G|Z}$  was based on an estimated range of 0.5-4.6 GBS cases per 10,000 ZIKV infections (21), which was approximated with a beta distribution ( $p_{G|Z} \sim \text{Beta}(5.9, 2.3 \times 10^4)$ ).

### *Blood bank indicator*

We used an adjustment factor ( $f_{BB}$ ) to account for the duration of test positivity and the proportion of infected individuals excluded from donating blood because they were symptomatic at the time of donation (16). Assuming weekly testing, the equations and distributions of *Chevalier et al.* were used to approximate a prior distribution for this factor:

$$f_{BB} = 7 * D / (p_A * V + (1 - p_A)).$$

We sampled from distributions of each component,  $p_A$ , the proportion of asymptomatic infections (5,6,11);  $V$ , the duration of viremia (23); and  $D$ , the incubation period (23,24), to estimate a Gamma prior for  $f_{BB}$ :  $f_{BB} \sim \text{Gamma}(6.7, 7.6)$ .

### *Analysis of priors*

We examined the effect of different prior distribution assumptions on model posteriors for the probability of a suspected case being reported ( $p_{S|Z}$ ), the probability of acquiring GBS if ZIKV infected ( $p_{G|Z}$ ), and the relative incidence of ZIKV in the general population compared to positivity in blood donor ( $f_{BB}$ ) parameters. We assessed the models with three alternative types of prior variance: informative (as described above), informative with doubled standard deviation in the prior and naïve (uniform or flattened). Prior distributions for each parameter under each assumption are available in Supplemental Table 1.

### *Model fitting*

We fitted epidemic models using a Markov chain Monte Carlo (MCMC) Bayesian framework to estimate incident ZIKV infections and 95% credible intervals [95% CrI] from the three surveillance indicators individually and using all three indicators combined. For each indicator model, we performed 1,001,000 iterations for three chains, and discarded the initial 1,000 iterations as the burn-in period. We evaluated convergence using the Gelman-Rubin diagnostic (25) and thinned the output using every 1,000th sample to obtain 1,000 effectively uncorrelated simulations per chain. For the MCMC simulations, we used the rstan package version 2.19.3 of Stan (version 2.18.0-1, Stan Development Team, <http://mc-stan.org>), and coda version 0.19-2 (26) package in R version 3.3.2 (<https://www.R-project.org/>).

## Results

### *Estimated infections based on individual surveillance indicators*

In 2016, 65,820 suspected arboviral disease cases, 175 suspected GBS cases, and 360 ZIKV-positive blood donors (out of 54,588 tested), were reported in Puerto Rico (Figure 1A-C). We used these data to estimate the weekly and cumulative ZIKV infection incidence in 2016 for each surveillance indicator independently (Figure 1D). Weekly infections estimated from suspected arbovirus cases and suspected GBS cases had similar trends over time, peaking in August 2016 and declining thereafter. Using the blood bank data, estimated ZIKV incidence peaked in June followed by high incidence through August, and declined afterwards. For cumulative incidence estimates based on each of the three indicators, the median estimate was lowest using suspected GBS cases (880,000 infections) and highest using blood bank testing data (960,000 infections), with substantial overlap of credible intervals (Table 2). Estimates based on suspected arbovirus cases had the lowest uncertainty (95% Credible Interval (95% CrI): 630,000 to 1,200,000) and estimates based on suspected GBS cases had the highest (95% CrI: 420,000 to 1,300,000). The estimated proportion of the population infected during the outbreak was similar across the three indicators, with 27% (95% CrI: 19% to 35%) infected using suspected arbovirus case indicator, 26% (95% CrI: 12% to 38%) using the suspected GBS indicator, and 28% (95% CrI: 19% to 37%) using the blood bank indicator.

### *Estimated infections with combined surveillance indicators*

Using a combination of the three surveillance indicators, infections peaked between August and September 2016 (Figure 1E), reflecting the combined peaks in incidence from the three indicators. We estimated that the total incident ZIKV infections was most likely between 810,000 and 970,000 infections (50% CrI, 95% CrI: 660,000 to 1,100,000) (Table 2), corresponding to 24% to 29% (50% CrI, 95% CrI: 19% to 33%) of the total population. These estimates correspond well to an *a priori* triangular probability distribution used to anticipate resource needs during the epidemic (Figure 2) (14,27). The combined estimates had reduced uncertainty compared to that triangle distribution and each independent estimate based on the individual surveillance indicators.



### *Prior and posterior parameter distributions*

For each of the estimates reported above, we used a model with an informed set of prior parameter distributions. However, we also compared these estimates to those with less informed priors (increased variance) and naïve priors (Figure 3A). Increased uncertainty in the priors resulted in similar median estimates for infections throughout 2016, but increased uncertainty especially for the lower bound of the credible intervals (Table 3).

For the three prior distributions, the transmission parameter ( $\beta_0$ ) converged on a similar value regardless of the surveillance data used. As expected, the less informative priors led to higher uncertainty in the posterior distributions for the outcome parameters, particularly for the individual models in which additional data are not available to inform the estimates (Figure 3B). The most notable effect was seen for the suspected GBS surveillance model. In this case, the posterior baseline GBS risk ( $p_{G0}$ ) was slightly higher and the ZIKV-specific infection risk ( $p_{G|Z}$ ) was highly uncertain, indicating a lack of sufficient information to distinguish between the two components of GBS risk. On the other hand, the combined model was able to resolve all parameters regardless of the assumed prior variance. However, parameters using naïve priors still had more uncertainty in their posterior distributions (Figure 3C).

### *Posterior estimates over time*

Over the progression of the outbreak, the parameter posteriors evolved over time as more data became available for each surveillance indicator (Figure 3D). For the first 4 weeks of 2016, the posterior estimates for individual parameters largely reflected the priors. However, by 8 weeks the posteriors started to shift, narrow, and stabilize.

We observed similar trends when assessing how the incidence estimates of the combined model changed over 4-week increments. When incorporating new data for each surveillance indicator into the combined model over the course of the 2016 epidemic, the incidence estimates had the largest uncertainty in the earliest weeks of the outbreak (Supplemental Figure 1). For each 4-week estimation, the end-of-year estimate based on the full dataset fell within the uncertainty bounds.

## Discussion

Estimates of the true burden of infection for ZIKV, like other pathogens, is challenging because many infections are inapparent; apparent infections are not always recognized, confirmed or reported; and disease surveillance systems for collecting case data are highly varied. Here, we developed an epidemic model, applied within a Bayesian modeling framework, to estimate ZIKV incidence in Puerto Rico using three separate infection indicators available from multiple surveillance systems and assumptions about detection probabilities for each system. Our approach utilizes different data sources to increase the precision of infection estimates over time and may further reduce bias by accounting for inherent surveillance biases based on the probability of detection. Using this framework, we estimated that ZIKV infections occurred in roughly a quarter of the population, resulting in 890,000 total incident ZIKV infections in Puerto Rico in 2016, translating to an average of 36 to 44 new infections per 10,000 people per week. The peak of the epidemic occurred during the week of August 15, 2016 (i.e. week 33), when an estimated 120-140 weekly incident infections occurred per 10,000 people.

We estimated that 19-33% of the population had symptomatic and asymptomatic ZIKV infections in Puerto Rico in 2016, a much higher proportion infected compared to the reported 36,316 confirmed cases (approximately 1% of the population) (3). This estimate was similar to estimates for other arboviral outbreaks in Puerto Rico and to ZIKV estimates generated using other approaches. Community-level dengue seroprevalence studies conducted after emerging outbreaks indicated infection rates of approximately 47% (range: 8-79%, 1969) (28) and 30% (range: 22-45%, 1982) (29). The 2014-2015 chikungunya epidemic resulted in 23.5% seropositivity among 1,031 blood donors in 2015 (30), and for communities participating in a chikungunya vector control study, seroprevalence was 23% (intervention) and 45% (non-intervention) (31). During the ZIKV epidemic, a household-based cluster investigation of 367 participants from 19 clusters found that ZIKV seroprevalence ranged from 0–57% (4). An estimate based on blood donors alone was 21% (95% confidence interval: 18–24%) (32) while an integrated approach similar to ours estimated an infection rate of 32% (95% CrI: 29-35%)(33).

In contrast to our analysis, studies from other islands found substantially higher seroprevalence estimates, including 73% in Yap in 2007 (5), 49% in French Polynesia in 2013–2014 (10), and 42%–50% in Martinique in 2015 (7,34). The differences in the estimated underlying infection burdens may be in part due to heterogeneity in exposure to infection even within an island population, as seen in early dengue serosurveys in Puerto Rico and municipality-level estimates for ZIKV infection rates (33). Invasion patterns, but nature drive some spatial heterogeneity which can influence longer-term transmission dynamics, since areas with high immunity may limit transmission to areas with low immunity. Other factors that may contribute to differing levels of immunity include human mobility, underlying socioeconomics, and cross-immunity from other arboviruses, and all warrant further examination. The estimated 2016 Puerto Rico infection rate is well below most model estimates for how large a Zika epidemic would be if the population was perfectly mixed (18).

The framework developed here offers advantages beyond the estimation of epidemic size. In practice, we used the model to actively inform situational awareness beginning in August 2016. From the analysis of performance over time, the model could have provided useful information even earlier in the epidemic, despite limited availability of information about the outcome probabilities. Informative priors were available early in the year for baseline GBS risk and for the blood bank reporting factor. While specification of informative prior distributions for the reporting of ZIKV cases and suspected GBS cases would have been more challenging, they would not have been completely naïve. Critically, combining indicators lessened the need for strong prior information for any single indicator. Assuming less precise priors had some effect on outcome precision for all models but had the least effect when indicators were combined, except in the case of completely naïve priors. Similar observations were reported by *Andronico et al.* (12), where examination of different prior assumptions did not strongly affect parameter estimates.

Our approach had some limitations. One of our key indicators, suspected arboviral cases, uses a broad case definition to capture all potential symptomatic cases, and likely includes cases caused by other circulating arboviruses, such as dengue and chikungunya viruses. However, these non-ZIKV cases presumably have had little impact on our indicator, as dengue and chikungunya were rare in Puerto Rico

in 2016 (3). We expect if there had been more dengue and chikungunya cases, our framework would improve differentiation from ZIKV cases from other arboviral cases. In general, our prior assumptions were based on data available at the time. If these early data contain hidden bias, resulting estimates could also be biased. Our results suggest that being conservative with respect to precision does not sacrifice substantial precision in the results, in part due to using multiple indicators. Our framework did not allow for substantial transmission variation beyond some week-to-week variability and reduced incidence due to acquired immunity in those already infected. In practice, transmission likely changes seasonally and can be influenced by vector control measures. In this case, with the introduction of a completely novel virus and the lack of large-scale effective vector control measures, we expect there would have been little change in transmissibility throughout the year, but that may not be appropriate in general. Seasonality and other changes in transmissibility can be incorporated into future versions of the framework.

Reported cases generally underestimate the true number of incident infections occurring in an epidemic, since they capture only recognized and reported clinical infections. However, multiple imperfect indicators provide the opportunity to estimate the underlying incidence of infection, utilizing multiple complementary indicators: (1) a broad non-specific indicator with relatively high counts (suspect arboviral disease cases); (2) an indicator of a severe outcome that is rare, but likely to have high reporting fidelity (GBS); and (3) an age- and geography-biased sample of infection prevalence (blood donors). Each indicator offers a unique but biased insight into the progression of the epidemic, capturing different case subgroups including different age-groups, and, when combined, can provide critical situational awareness about the progression of the epidemic.

The approach described here can estimate how many people have been infected in near real-time or identify changes in the trajectory of incidence across various indicators. It is also useful for post-hoc analysis to understand what the impact may have been on the population-level and whether more transmission may be expected. With 19-33% of the population infected in 2016, ZIKV transmission should be much more limited but still possible in Puerto Rico, particularly in areas that may have experienced lower infection rates. These insights are critical both for preparedness for and response to

future epidemics, and this modeling approach is applicable to future Zika epidemics as well as epidemics of other pathogens. For instance, using reported influenza-like-illness (ILI)-associated hospitalizations, outpatient ILI visits and reported laboratory-confirmed specimens each have limitations as individual indicators of the incidence of influenza infection, but when combined they may best approximate the incidence of influenza infection. Approaches like the one we present here provide a tool to incorporate these diverse data and the uncertainties in them to generate timely estimates of incident infection and inform response and control efforts.

## Acknowledgements

The findings in this article are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention or the U.S. Public Health Service.

## References

1. Fauci AS, Morens DM. Zika Virus in the Americas--Yet Another Arbovirus Threat. *N Engl J Med*. 2016 Feb;374(7):601–4.
2. Thomas DL, Sharp TM, Torres J, Armstrong PA, Munoz-Jordan J, Ryff KR, et al. Local Transmission of Zika Virus--Puerto Rico, November 23, 2015-January 28, 2016. *MMWR Morb Mortal Wkly Rep*. 2016 Feb;65(6):154–8.
3. Sharp TM, Quandelacy TM, Adams LE, Aponte JT, Lozier MJ, Ryff K, et al. Epidemiologic and spatiotemporal trends of Zika Virus disease during the 2016 epidemic in Puerto Rico. *PLoS Negl Trop Dis* [Internet]. 2020 Sep 21;14(9):e0008532. Available from: <https://doi.org/10.1371/journal.pntd.0008532>
4. Lozier MJ, Burke RM, Lopez J, Acevedo V, Amador M, Read JS, et al. Differences in Prevalence of Symptomatic Zika Virus Infection, by Age and Sex-Puerto Rico, 2016. *J Infect Dis*. 2018 May;217(11):1678–89.

5. Duffy MR, Chen T-H, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med*. 2009 Jun;360(24):2536–43.
6. Musso D, Nhan T, Robin E, Roche C, Bierlaire D, Zisou K, et al. Potential for Zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia, November 2013 to February 2014. *Euro Surveill Bull Eur sur les Mal Transm = Eur Commun Dis Bull*. 2014 Apr;19(14).
7. Gallian P, Cabié A, Richard P, Paturel L, Charrel RN, Pastorino B, et al. Zika virus in asymptomatic blood donors in Martinique. Vol. 129, *Blood*. United States; 2017. p. 263–6.
8. Krauer F, Riesen M, Reveiz L, Oladapo OT, Martínez-Vega R, Porgo T V, et al. Zika Virus Infection as a Cause of Congenital Brain Abnormalities and Guillain-Barré Syndrome: Systematic Review. *PLoS Med*. 2017 Jan;14(1):e1002203.
9. Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR. Zika Virus and Birth Defects--Reviewing the Evidence for Causality. *N Engl J Med*. 2016 May;374(20):1981–7.
10. Aubry M, Teissier A, Huart M, Merceron S, Vanhomwegen J, Roche C, et al. Zika Virus Seroprevalence, French Polynesia, 2014-2015. *Emerg Infect Dis*. 2017 Apr;23(4):669–72.
11. Mitchell PK, Mier-Y-Teran-Romero L, Biggerstaff BJ, Delorey MJ, Aubry M, Cao-Lormeau V-M, et al. Reassessing Serosurvey-Based Estimates of the Symptomatic Proportion of Zika Virus Infections. *Am J Epidemiol*. 2019 Jan;188(1):206–13.
12. Andronico A, Dorléans F, Fergé J-L, Salje H, Ghawché F, Signate A, et al. Real-Time Assessment of Health-Care Requirements During the Zika Virus Epidemic in Martinique. *Am J Epidemiol*. 2017 Nov;186(10):1194–203.
13. Sharp TM, Fischer M, Muñoz-Jordán JL, Paz-Bailey G, Staples JE, Gregory CJ, et al. Dengue and Zika Virus Diagnostic Testing for Patients with a Clinically Compatible Illness and Risk for Infection with Both Viruses. *MMWR Recomm reports Morb Mortal Wkly report Recomm reports*. 2019 Jun;68(1):1–10.

14. Dirlikov E, Kniss K, Major C, Thomas D, Virgen CA, Mayshack M, et al. Guillain-Barré Syndrome and Healthcare Needs during Zika Virus Transmission, Puerto Rico, 2016. *Emerg Infect Dis.* 2017 Jan;23(1):134–6.
15. Dirlikov E, Ryff KR, Torres-Aponte J, Thomas DL, Perez-Padilla J, Munoz-Jordan J, et al. Update: Ongoing Zika Virus Transmission - Puerto Rico, November 1, 2015-April 14, 2016. *MMWR Morb Mortal Wkly Rep.* 2016 May;65(17):451–5.
16. Chevalier MS, Biggerstaff BJ, Basavaraju S V, Ocfemia MCB, Alsina JO, Climent-Peris C, et al. Use of Blood Donor Screening Data to Estimate Zika Virus Incidence, Puerto Rico, April-August 2016. *Emerg Infect Dis.* 2017 May;23(5).
17. US Census Bureau. Quick Facts - Puerto Rico [Internet]. 2016. Available from: <https://www.census.gov/quickfacts/PR>
18. Keegan LT, Lessler J, Johansson MA. Quantifying Zika: Advancing the Epidemiology of Zika With Quantitative Models. *J Infect Dis.* 2017 Dec;216(suppl\_10):S884–90.
19. Perkins TA, Metcalf CJE, Grenfell BT, Tatem AJ. Estimating drivers of autochthonous transmission of chikungunya virus in its invasion of the americas. *PLoS Curr.* 2015 Feb;7.
20. Ferguson NM, Cucunubá ZM, Dorigatti I, Nedjati-Gilani GL, Donnelly CA, Basáñez M-G, et al. EPIDEMIOLOGY. Countering the Zika epidemic in Latin America. *Science.* 2016 Jul;353(6297):353–4.
21. Mier-Y-Teran-Romero L, Delorey MJ, Sejvar JJ, Johansson MA. Guillain-Barré syndrome risk among individuals infected with Zika virus: a multi-country assessment. *BMC Med.* 2018 May;16(1):67.
22. Sejvar JJ, Baughman AL, Wise M, Morgan OW. Population incidence of Guillain-Barré syndrome: a systematic review and meta-analysis. *Neuroepidemiology.* 2011;36(2):123–33.
23. Lessler J, Ott CT, Carcelen AC, Konikoff JM, Williamson J, Bi Q, et al. Times to key events in Zika virus infection and implications for blood donation: a systematic review. *Bull World Health Organ.* 2016 Nov;94(11):841–9.

24. Lessler J, Chaisson LH, Kucirka LM, Bi Q, Grantz K, Salje H, et al. Assessing the global threat from Zika virus. *Science* (80- ) [Internet]. 2016 Aug 11;353(6300). Available from: <http://science.sciencemag.org/content/353/6300/aaf8160.abstract>
25. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci*. 1992;7(4):457–511.
26. Vines MP and NB and KC and K. CODA: Convergence Diagnosis and Output Analysis for MCMC [Internet]. *R News*. 2006 [cited 2020 Jul 8]. p. 7–11. Available from: <https://journal.r-project.org/archive/>
27. Ellington SR, Devine O, Bertolli J, Martinez Quiñones A, Shapiro-Mendoza CK, Perez-Padilla J, et al. Estimating the Number of Pregnant Women Infected With Zika Virus and Expected Infants With Microcephaly Following the Zika Virus Outbreak in Puerto Rico, 2016. *JAMA Pediatr*. 2016 Oct;170(10):940–5.
28. Likosky WH, Calisher CH, Michelson AL, Correa-Coronas R, Henderson BE, Feldman RA. An epidemiologic study of dengue type 2 in Puerto Rico, 1969. *Am J Epidemiol*. 1973 Apr;97(4):264–75.
29. Waterman SH, Novak RJ, Sather GE, Bailey RE, Rios I, Gubler DJ. Dengue transmission in two Puerto Rican communities in 1982. *Am J Trop Med Hyg*. 1985 May;34(3):625–32.
30. Simmons G, Brès V, Lu K, Liss NM, Brambilla DJ, Ryff KR, et al. High Incidence of Chikungunya Virus and Frequency of Viremic Blood Donations during Epidemic, Puerto Rico, USA, 2014. *Emerg Infect Dis*. 2016 Jul;22(7):1221–8.
31. Lorenzi OD, Major C, Acevedo V, Perez-Padilla J, Rivera A, Biggerstaff BJ, et al. Reduced Incidence of Chikungunya Virus Infection in Communities with Ongoing Aedes Aegypti Mosquito Trap Intervention Studies - Salinas and Guayama, Puerto Rico, November 2015-February 2016. *MMWR Morb Mortal Wkly Rep*. 2016 May;65(18):479–80.
32. Williamson PC, Biggerstaff BJ, Simmons G, Stone M, Winkelman V, Latoni G, et al. Evolving viral and serological stages of Zika virus RNA-positive blood donors and estimation of incidence



- of infection during the 2016 Puerto Rican Zika epidemic: an observational cohort study. *Lancet Infect Dis.* 2020 Jul;
33. Moore SM, Oidtman RJ, Soda J, Siraj AS, Reiner RC, Barker CM, et al. Leveraging multiple data types to estimate the true size of the Zika epidemic in the Americas. *medRxiv* [Internet]. 2019 Jan 1;19002865. Available from: <http://medrxiv.org/content/early/2019/08/22/19002865.abstract>
34. Cousien A, Abel S, Monthieux A, Andronico A, Calmont I, Cervantes M, et al. Assessing Zika Virus Transmission Within Households During an Outbreak in Martinique, 2015–2016. *Am J Epidemiol* [Internet]. 2019 Apr 17;188(7):1389–96. Available from: <https://doi.org/10.1093/aje/kwz091>
35. Cao-Lormeau V-M, Blake A, Mons S, Lastère S, Roche C, Vanhomwegen J, et al. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet* [Internet]. 2016;387(10027):1531–9. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0140673616005626>
36. MACNAMARA FN. Zika virus: a report on three cases of human infection during an epidemic of jaundice in Nigeria. *Trans R Soc Trop Med Hyg.* 1954 Mar;48(2):139–45.
37. HAMMON WM, SCHRACK WDJ, SATHER GE. Serological survey for a arthropod-borne virus infections in the Philippines. *Am J Trop Med Hyg.* 1958 May;7(3):323–8.
38. Cauchemez S, Besnard M, Bompard P, Dub T, Guillemette-Artur P, Eyrolle-Guignot D, et al. Association between Zika virus and microcephaly in French Polynesia, 2013–2015: a retrospective study. *Lancet* [Internet]. 2016 May 21;387(10033):2125–32. Available from: [https://doi.org/10.1016/S0140-6736\(16\)00651-6](https://doi.org/10.1016/S0140-6736(16)00651-6)
39. DICK GWA, KITCHEN SF, HADDOW AJ. Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg.* 1952 Sep;46(5):509–20.
40. Robin Y, Mouchet J. [Serological and entomological study on yellow fever in Sierra Leone]. *Bull Soc Pathol Exot Filiales.* 1975;68(3):249–58.
41. SMITHBURN KC. Neutralizing antibodies against arthropod-borne viruses in the sera of long-

- time residents of Malaya and Borneo. *Am J Hyg.* 1954 Mar;59(2):157–63.
42. SMITHBURN KC. Neutralizing antibodies against certain recently isolated viruses in the sera of human beings residing in East Africa. *J Immunol.* 1952 Aug;69(2):223–34.
  43. POND WL. ARTHROPOD-BORNE VIRUS ANTIBODIES IN SERA FROM RESIDENTS OF SOUTH-EAST ASIA. *Trans R Soc Trop Med Hyg.* 1963 Sep;57:364–71.

## Tables

**Table 1. Model parameters and prior distributions**

Parameter	Description	Mean	95% CI	Distribution	Reference
Population (N)	2016 population for Puerto Rico	3,400,000	-	-	(17)
Suspect Zika probability ( $p_{S Z}$ )	Probability of ZIKV infection becoming a suspect case	0.11	0.01, 0.24	Beta	(18)
Viremic detection (D)	Period of detecting Zika virus RNA in blood (days)	10	3, 18	Weibull	(16)
Asymptomatic proportion ( $p_A$ )	Proportion of ZIKV infections that are asymptomatic	0.68		Beta	(5,6,16)
GBS probability ( $p_{G Z}$ )	Probability of GBS given ZIKV infection	1.39/100,000 per week		Beta	(35)
Global baseline GBS risk ( $p_{G0}$ )	Weekly GBS risk due to other causes	1.89/100,00 per week		Beta	(21)

Abbreviations: CI, Confidence interval; GBS, Guillain-Barré Syndrome; ZIKV, Zika virus

**Table 2. Estimated Zika virus infections and 95% credible intervals [CrI] using three surveillance indicators**

<b>Surveillance Indicator</b>	<b>Cumulative ZIKV infections [95% CrI]</b>	<b>Proportion of ZIKV infections [95% CrI]</b>	<b>Median incident infections per 10,000 [95% CrI]</b>	<b>Peak weekly incident infections per 10,000 people [95% CrI]</b>	<b>Week of peak incidence [95% CrI]</b>
<b>Suspected arboviral cases</b>	900,000 [630,000, 1,200,000]	27% [19%, 35%]	41 [28, 54]	130 [93, 180]	33 [32, 34]
<b>Suspected ZIKV- associated GBS cases</b>	880,000 [420,000, 1,300,000]	26% [12%, 38%]	39 [16, 56]	130 [66, 240]	32 [28, 38]
<b>Blood bank data</b>	960,000 [650,000, 1,300,000]	28% [19%, 37%]	35 [22, 48]	150 [99, 220]	28 [24, 34]
<b>Combined indicators</b>	890,000 [660,000, 1,100,000]	26% [19%, 33%]	40 [30, 52]	130 [97, 170]	33 [22, 34]

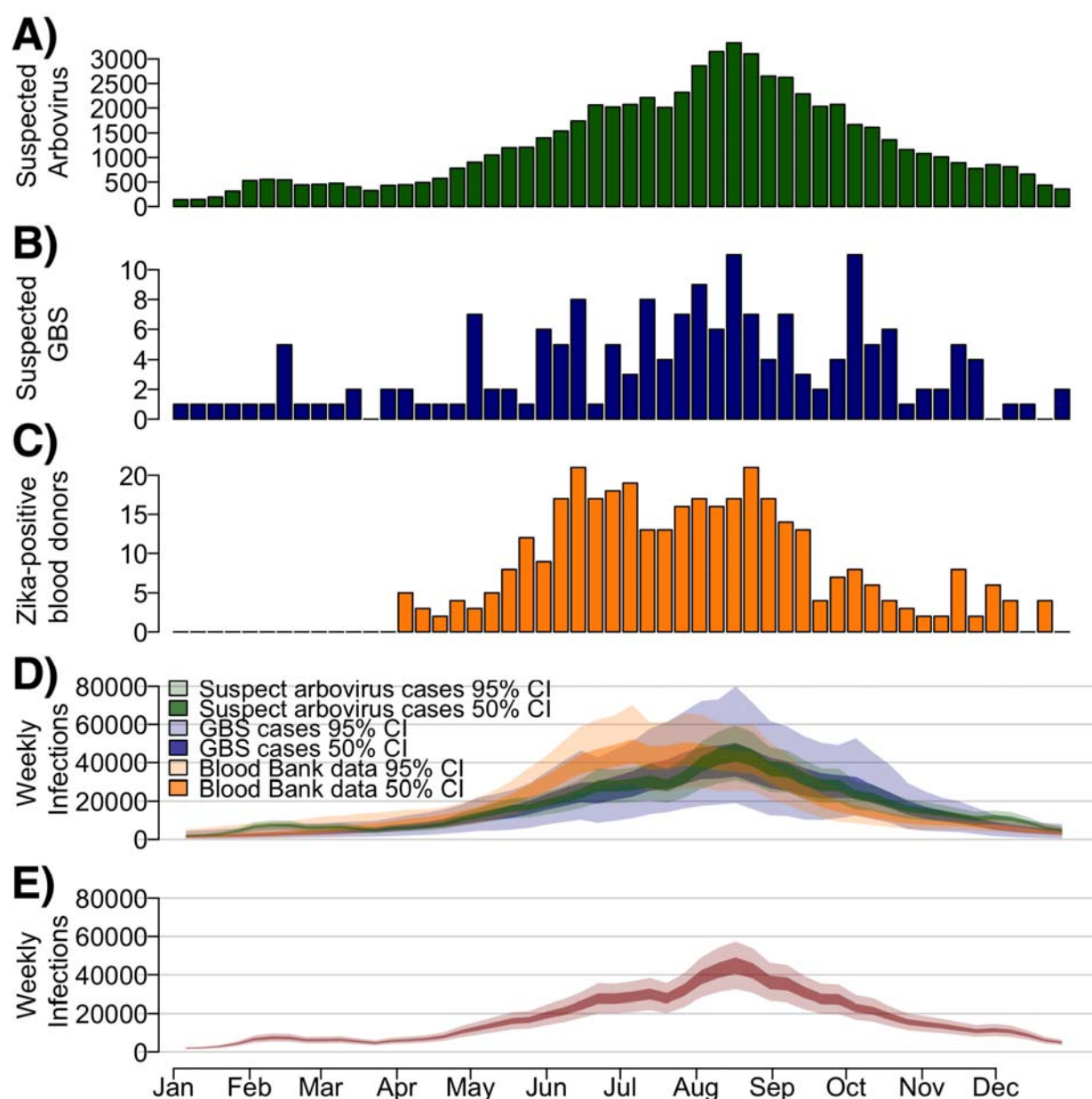
Median and 95% CrIs are shown from posterior distributions for each surveillance indicator. Peak week refers to the calendar week of the year associated with peak incidence during the epidemic. Abbreviations: CrI, credible interval; ZIKV, Zika virus

**Table 3. Estimated cumulative Zika virus (ZIKV) infections from models using informative, naïve, and increased variances for prior distributions, Puerto Rico, 2016**

	Assumed Prior Parameter Distribution					
	Informative		Increased variance		Naïve	
Surveillance Indicator	Cumulative ZIKV infections [95% CrI]	Proportion of ZIKV infections [95% CrI]	Cumulative ZIKV infections [95% CrI]	Proportion of ZIKV infections [95% CrI]	Cumulative ZIKV infections [95% CrI]	Proportion of ZIKV infections [95% CrI]
<b>Suspected arboviral cases</b>	900,000 [630,000, 1,200,000]	27% [19%, 35%]	910,000 [590,000, 1,200,000]	27% [17%, 34%]	910,000 [610,000, 1,200,000]	27% [18%, 24%]
<b>Suspected ZIKV-associated GBS cases</b>	880,000 [420,000, 1,300,000]	26% [12%, 38%]	900,000 [330,000, 1,400,000]	31% [10%, 42%]	810,000 [380, 1,400,000]	24% [0%, 42%]
<b>Blood bank data</b>	960,000 [650,000, 1,300,000]	28% [19%, 37%]	900,000 [530,000, 1,200,000]	26% [16%, 35%]	780,000 [100,000, 1,200,000]	23% [3%, 34%]
<b>Combined indicators</b>	890,000 [660,000, 1,100,000]	26% [19%, 33%]	870,000 [600,000, 1,110,000]	26% [18%, 33%]	860,000 [470,000, 1,110,000]	25% [14%, 33%]

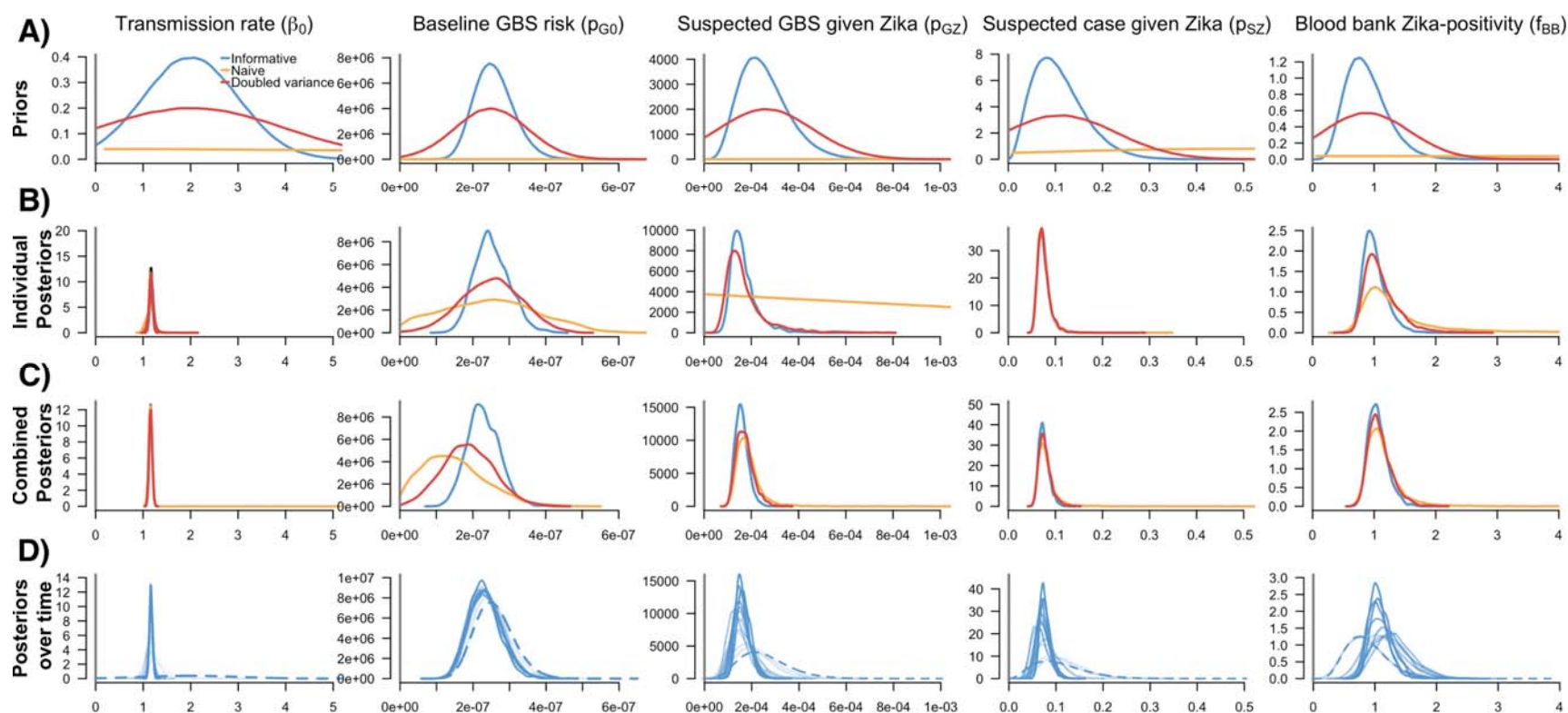
Median and 95% CrIs are shown from posterior distributions for each surveillance indicator. Abbreviations: CrI, credible interval; Guillain-Barré Syndrome; ZIKV, Zika virus

## Figures



**Figure 1. Suspected arbovirus cases, suspected ZIKV-associated Guillain-Barré Syndrome (GBS) cases, ZIKV-positive blood donors, and estimated weekly Zika infections during the 2016 outbreak in Puerto Rico.** A) Number of suspected arbovirus cases reported (green). B) Number of suspected ZIKV-associated GBS cases reported. C) Number of ZIKV-positive blood donors identified from blood donor screening. D) Estimated weekly infections using each indicator model separately. Colors refer to





**Figure 3. Prior and posterior parameter distributions from individual indicator models, the combined indicator model, and the combined model over time.** A) Prior distributions of five model parameters. Color lines refer to assessed variance assumptions of prior distributions in sensitivity analyses. Final individual and combined indicator models used informative priors. B) Posterior distributions of model parameters from individual indicator models. The dashed lines for the Beta parameter refer to the posterior parameter distributions from three models. C) Posterior distributions of model parameters from the combined model. D) Posterior distributions over time (i.e. four-week increments from the end of



January 2016 to the end of December 2016) from the combined model using informative priors. Dashed lines refer to the informative priors for each model parameter. Darker transparency of the lines refers to each the posterior distribution from each 4-week increment over time.