

RESEARCH ARTICLE

Estimating under-observation and the full size of the 2016 Zika epidemic in Rio de Janeiro

Marcio M. Bastos^{✉*}, Flávio Codeço Coelho[✉]

Escola de Matemática Aplicada, Fundação Getúlio Vargas, Rio de Janeiro, RJ, Brazil

✉ These authors contributed equally to this work.

* marciomacielbastos@hotmail.com



OPEN ACCESS

Citation: Bastos MM, Coelho FC (2018) Estimating under-observation and the full size of the 2016 Zika epidemic in Rio de Janeiro. PLoS ONE 13(10): e0205001. <https://doi.org/10.1371/journal.pone.0205001>

Editor: Jeffrey Shaman, Columbia University, UNITED STATES

Received: February 3, 2018

Accepted: September 18, 2018

Published: October 12, 2018

Copyright: © 2018 Bastos, Coelho. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are available from: <https://github.com/AlertaDengue/publication-data>.

Funding: This project has received financial support in the form of a scholarship to Marcio Bastos from Fundação Getúlio Vargas, and equipment in the form of a TESLA GPU card from NVIDIA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The funder NVIDIA did not require any recompense in return for the provided

Abstract

The 2015–16 Zika epidemic spread quickly from north to south in Brazil. Two striking features were the much higher incidence in young adult women due to sexual transmission, and the serious congenital malformations and miscarriages associated to Zika infection in pregnant women. In this paper we use case reporting data along with live-birth records to reconstruct the full size of the epidemic through a Bayesian probabilistic graph model representing the Zika transmission probabilities of observation (case reporting) and of birth loss (through miscarriage or abortion). We find that the probability of observing (reporting) a Zika case is different between men and women and ranges between 10 to 13%. From these estimates we reconstruct the full size of the Zika epidemic in Rio de Janeiro in 2015–16.

Introduction

In the summer of 2015–16, the city of Rio de Janeiro suffered from a epidemic of Zika which had been introduced months earlier in the northeast of the country. The quick spread of Zika and the particular way in which it attacked the female population [1] led to serious public health consequences. Rio de Janeiro had one of the country's largest incidences of microcephaly associated with Zika. Another important impact was the marked number of lost pregnancies associated with Zika [2].

Despite its serious implications, the clinical diagnosis of Zika in Rio was not easy because of the combination of mild symptoms and the difficulty of distinguishing it from the simultaneous outbreaks of Dengue and Chikungunya taking place in the city.

Nevertheless, the speed with which the epidemic covered the entire country seems to point to a high transmissibility and a low observation rate [3, 4]. Here we define observation as a clinically diagnosed patient. Due the high state of alert of the health system, a fairly high notification rate (given a clinical diagnosis) was maintained throughout the epidemic.

Estimating the parameters that govern its transmission, specially the *basic reproduction number* (\mathcal{R}_0) [5], is important to comprehend its dynamics, because it gives a metric to compare the spread of the disease in different areas and to help automatically detect patterns, leading to more accurate forecasts and better prevention, therapy and control [6].

hardware. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

The Zika outbreak has already been studied in other regions with different probabilistic approaches such as the study of the transmission dynamics during the epidemic in 2013–2014 in French Polynesia [7] and the estimation of the Reproduction Number and the role of the sexual transmission of the epidemic in 2015 in Barranquilla, Colombia [8].

As one of our goals in this paper is to estimate the under-reporting rate, we use case reporting data which also give us reliable estimates about the relative incidences by age and gender. The complete birth record in Rio de Janeiro, including the drop in births associated to female cases of Zika during the 2016 epidemic [2], helps us estimate the probability of observing a case of Zika, and consequently the full size of the epidemic in the city of Rio de Janeiro.

This paper makes two key contributions to the understanding of Zika: firstly, it provides a realistic estimate of the real magnitude of the Zika epidemic in Rio de Janeiro in the year 2016 and secondly it measures the drop in births caused by Zika [2] through a non-iterative methodology for the estimation of parameters of an SIR transmission model using Bayesian inference.

Methods

In order to estimate the global epidemiological parameters for Zika from the available data, we approximated the transmission dynamics by an SIR model. In this model, we assumed that the total population in the city is constant in the study period ($N = S(t) + I(t) + R(t), \forall t$) and is divided into the susceptible, infectious and removed classes. The dynamics between these classes are governed by the Eqs (1), (2) and (2). This approximation of vector-borne transmission by an SIR model has been applied successfully before [9] and is justifiable due to the lack of specific data on the vector.

$$\frac{dS}{dt} = \frac{-\beta IS}{N} \quad (1)$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

Although Zika is transmitted *via* both mosquitoes and sexual contact [1] the total transmissibility is represented by a single parameter β in the model used here. As the time period considered is short, a single recovery rate γ is also assumed corresponding mostly to the viremic period ($1/\gamma$), where the long term effects of sexual infectiousness in males are not likely to be relevant.

As the Reproductive Number (\mathcal{R}) is defined by (4), we used the analogous Eqs (5) and (6) to represent the dynamics between the classes.

$$\mathcal{R} = \frac{\beta}{\gamma} \quad (4)$$

$$\frac{dS}{dt} = \frac{-\gamma \mathcal{R} IS}{N} \quad (5)$$

$$\frac{dI}{dt} = \frac{\gamma \mathcal{R} IS}{N} - \gamma I \quad (6)$$

Because we had the opportunity to combine the evidence on Zika incidence in women in the reproductive age bracket with the number of lost births associated with Zika from the birth

records, we based the inference on the female epidemic (restricted to fertile age) and then derived the parameter for the total female and male epidemic from that of females in fertile age estimates.

It is important to mention that during the inference we did not simulate the dynamical model, but rather use its constraints in the bayesian inference model as detailed below.

In this paper we used the term observability most of the time, as we wanted to make a clear distinction between notified cases (necessarily observed) and those not notified because patients never sought medical attention, or were not recognized as a Zika case. We assumed the latter group to consist mostly of asymptomatic and very mild Zika cases, given the expected high ratio of asymptomatic to symptomatic cases [4, 10].

The data set

The data used in this paper comes from two public health information systems, *Sistema de Informação de Agravos e Notificação* (SINAN) and *Sistema de Informações sobre Nascidos Vivos* (SINASC).

From SINAN we extracted the Zika incidence ranging from the first epidemiological week of 2015 to the end of 2016. The registry for live births, SINASC, is a very well kept and complete registry, from which we extracted the total number live births recorded from 2012 to 2016. Both sets of data used had been fully anonymized prior to access by the authors.

In our study, we used official population estimates, calculated from the last census in 2010 for men ($N_m = 2,994,018$) and women ($N_f = 3,392,425$).

Data preparation. The incidence data extracted from the SINAN were aggregated weekly and separated by sex (\mathcal{D}_f and \mathcal{D}_m) and the subset of reports in fertile age (\mathcal{D}_{ffa} —between 14 and 48 years old) was selected from the female data (\mathcal{D}_f).

Splitting the incidence data in this way was important because we exploit the relation between the decrease in number of births (in the studied period) and the sexual transmission of Zika virus. In addition, we already know that the sexual transmission of the Zika virus is stronger from male to female [11]. That way, the \mathcal{D}_{ffa} dataset is the cornerstone on which we built the mathematical model for inferring the other fractions of populations, represented by their respective data stratus.

From the SINASC data, we define \mathcal{D}_{mz} . This time series is used as proxy for the weekly number of lost births due to Zika in Rio de Janeiro in 2016. Arboviroses transmitted by Aedes aegypti display marked seasonality, with the incidence peaking at the end of summer.

We restricted the inference to the data within the epidemic season. In order to delimit the epidemic season, we assume the incidence data is a poison process (Eq (7)) with two distinct intensities, a lower intensity corresponding to off-season and a higher average intensity corresponding to the epidemic season.

$$O_t \sim \text{Poisson}(\lambda_t) \quad (7)$$

The Poisson process intensity λ can be estimated from the data. We make the simplifying assumption that λ_t is constant in the high and low seasons, we have to estimate three values for λ (Eq (8)) and two additional parameters τ_1 and τ_2 corresponding to the transition times as showed in the random variable O (Eq (9)), where $\mathbb{1}_{(condition)}$ is the indicator function.

$$\lambda = \begin{cases} \lambda_1 & \text{if } t \leq \tau_1 \\ \lambda_2 & \text{if } \tau_1 \leq t \leq \tau_2 \\ \lambda_3 & \text{if } \tau_2 \leq t \end{cases} \quad (8)$$

$$O|\lambda, \tau \sim Poisson(\lambda = \mathbb{1}_{(week \leq \tau)} \cdot \lambda_i + \mathbb{1}_{(week > \tau)} \cdot \lambda_j) \quad (9)$$

That way, the epidemic season becomes bounded by the transition points τ_1 and τ_2 as shown in Fig 1. The inference problem can be depicted by the probabilistic graphical model of Fig 2. We applied a MCMC-based Bayesian inference procedure to this model to estimate its latent variables.

The prior distribution of τ (Eq (10)) is modeled as a discrete uniform on the interval $[0, \hat{n}]$, where \hat{n} is the number of weeks in \mathcal{D}_1 .

Since λ_k ($k \in \{1, 2, 3\}$) is the mean of a Poisson random variable, it must be positive definite. Thus, we choose exponential distributions to the priors, as shown in Eq (11). Also, the hyper-parameter in each λ_k was arbitrarily chosen to be equal to the inverse of the average of the count data (a usual procedure for exponential parameters).

The prior distributions of the latent variables are given below.

$$\tau \sim \mathcal{U}\{0, \hat{n}\} \quad (10)$$

$$\lambda_k \sim Exp(1/\bar{y}) \quad (11)$$

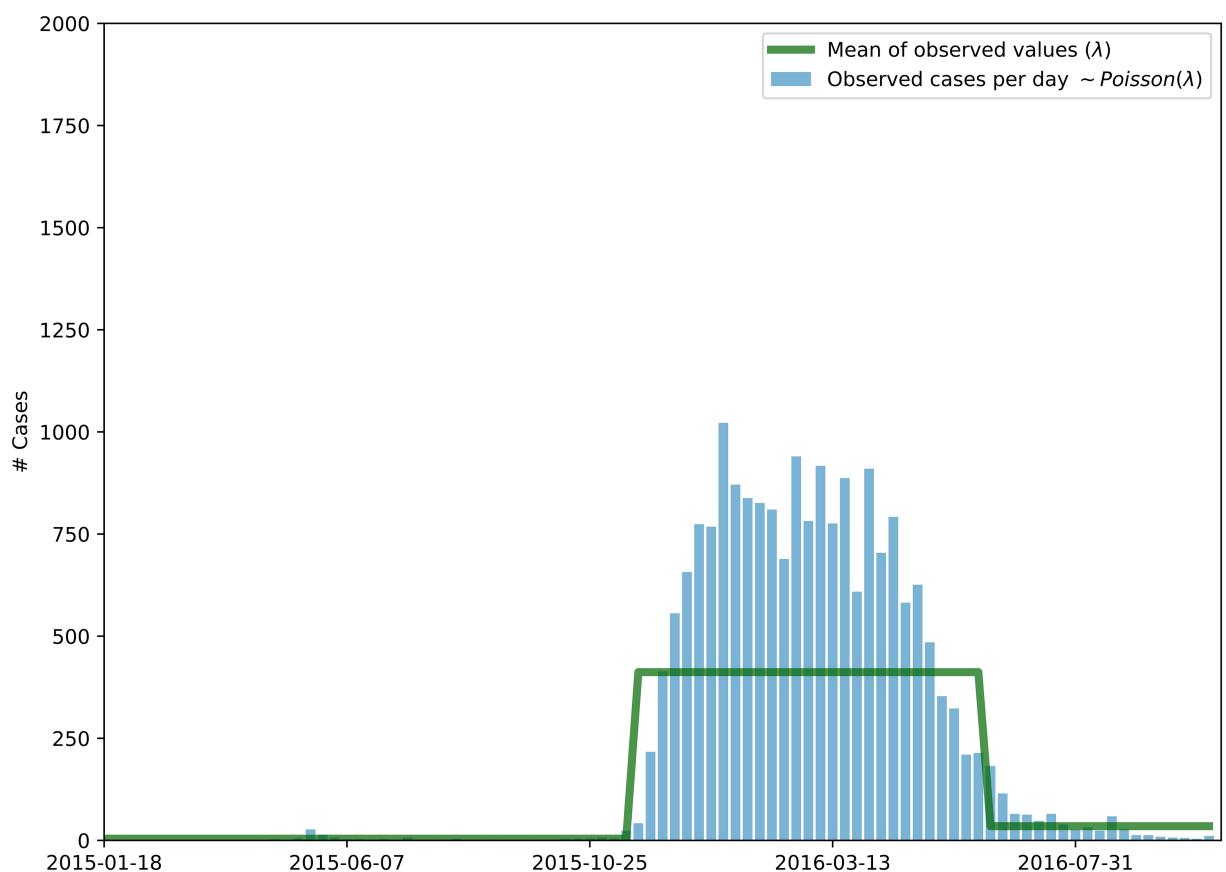


Fig 1. Period of interest. Given that the Zika's reports follow Poisson distribution, the green curve is the value of the parameter of the distribution in each week, the points of maximum in the interval bounds the period of significant infection, being thus the period that will be analyzed.

<https://doi.org/10.1371/journal.pone.0205001.g001>

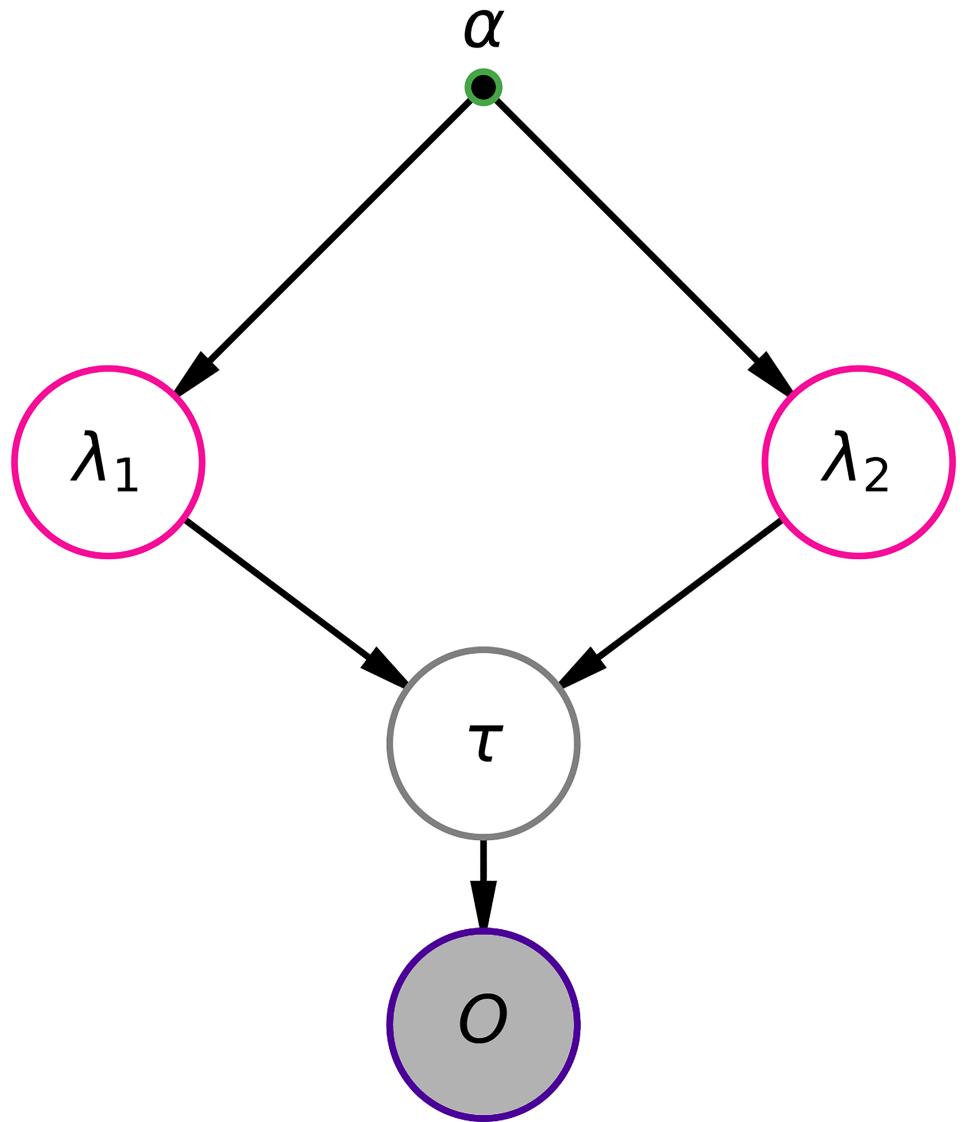


Fig 2. Bayesian network for estimating the time boundaries of the epidemic season. The gray filled circle means the observed data, the green border dots to represent a fixed parameter, the gray border circle indicate a *Uniformly* distributed variable, the pink border variable indicate *Exponentially* distributed variables and the purple halo means *Poisson* distribution. The arrows indicate probabilistic dependencies.

<https://doi.org/10.1371/journal.pone.0205001.g002>

The estimated bounds of the epidemic season, τ_1 and τ_2 , were used to temporally clip \mathcal{D}_1 . By retaining only case reports for females in fertile age, we obtained \mathcal{D}_{ffa} . After this we were left with $n = 29$ weeks of data to analyze.

The same time window was applied to the remaining case sets \mathcal{D}_f , \mathcal{D}_m and \mathcal{D} , respectively the total female cases, male cases and the total population cases.

We followed the procedure described in [2] to generate a weekly count of missing births or miscarriages (\mathcal{D}_{mz}).

Simply put, the procedure consists in subtracting the number of births in the i -th week of 2016 from the i -th week of 2015, as shown by the Eq (12), where B_j is the array of weekly

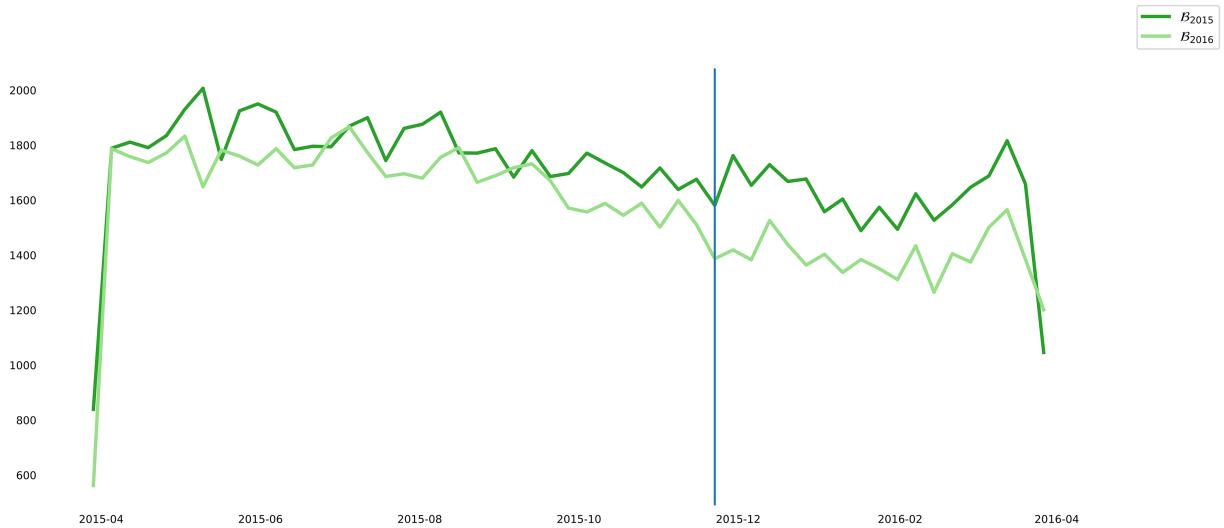


Fig 3. Overlapping of time series with births in 2015 and 2016. The vertical line marks the 40-th week.

<https://doi.org/10.1371/journal.pone.0205001.g003>

number of live births in the j -th year.

$$D'_{mz} = B_{2015} - B_{2016} \quad (12)$$

According to [2] the loss of births trails the Zika epidemic by about 40 weeks, as pictured in the Fig 3. Thus, D'_{mz} was shifted 40 weeks backward and bounded to the same time window as \mathcal{D}_{fia} . The weeks with no data in the bounded D'_{mz} were filled with 0 (assuming no losses), thus generating \mathcal{D}_{mz} .

Likelihoods

The case report data in each population fraction was modeled as a Poisson random variable ΔS_o (Eq (13)), whose parameter is the product of the incidence (ΔS) and the probability of observation (po).

Let pp be the probability of a woman (of fertile age) being pregnant and mz the probability of one of these women losing her baby due to Zika. Therefore, \mathcal{D}_{mz} was modeled as a Poisson random variable (M) with parameter $\Delta S \cdot pp \cdot mz$, as shown in Eq (14). This likelihood was used only in the female in fertile age model.

$$\Delta S_o \sim \text{Poisson}(\Delta S \cdot po) \quad (13)$$

$$M \sim \text{Poisson}(\Delta S \cdot pp \cdot mz) \quad (14)$$

Female in fertile age model

The first model we ran was just for female population in fertile age. The following priors were used.

We know from [7] that 11.5% (95%CI: 7.32–17.9%) of infections were reported overall in that study. Thus, assuming that the mildness of most Zika cases could lead to similar levels of underreporting in our data, we defined the prior probability of observation (po) to follow a

beta distribution as per Eq (15), where the parameters α_{po} and β_{po} were chosen so that the mean and the standard deviation agree with the mean and 95% credible interval previously reported.

As established in [2], there is a strong association between the reduction in the observed fertility in the period of analysis (2016) and the Zika epidemic in the same year. With this knowledge, we assumed that the percentage of fertility reduction caused by Zika (mz) follows a beta distribution whose parameters α_{mz} and β_{mz} imply a mean of 0.5 and 95% probability between 0.25 and 0.75. The distribution of mz is given by Eq (17).

The variable pp (Eq (16)) represents the probability of a woman in fertile age being pregnant. We choose a beta prior with mean equal to the ratio between the number of births in 2015 and the number of women in fertile age.

For the SIR model state variables, we proposed vector shaped priors since we fitted the entire series to data at once. We chose prior probabilities functions that fit the definition of each variable.

The parameter \mathcal{R}_t (Eq 19), which represents the ratio between the transmission rate of an individual and the period in which it remains infectious, was modeled by a vector-shaped random variable, with each element of the vector representing one week of the studied period. We allowed for its value to change from week to week to accommodate the effect of fluctuating environmental conditions not explicitly included in the model. Also, we used as prior a gamma distribution such that the parameters α_r and β_r follows the results of [7] in the the Moorea region (mean 4.8 and (95%CI: 3.2–8.4)).

We assumed that the infectivity period (γ^{-1}) was constant throughout the year and follows a gamma distribution with an average of 7 days and a standard deviation of 3.5. This choice was made to cover the infectious period and the life span of the mosquito (5 and 14 days respectively, as shown in [7]). This prior is summarized by the variable (18), the parameters α_γ and β_γ follow the restrictions given above.

$$po \sim Beta(\alpha_{po}, \beta_{po}) \quad (15)$$

$$pp \sim Beta(\alpha_{pp}, \beta_{pp}) \quad (16)$$

$$mz \sim Beta(\alpha_{mz}, \beta_{mz}) \quad (17)$$

$$\gamma^{-1} \sim Gamma(\alpha_\gamma, \beta_\gamma) \quad (18)$$

$$\mathcal{R}_t \sim Gamma(\alpha_r, \beta_r) \quad (19)$$

With no prior information about the weekly total number of infected persons (I), we chose a more flexible and decentralized model to this variable, capable of, using a Hamiltonian-based sampler as No-U-Turn Sampler (NUTS) (described in [12]), to efficiently span the sample space to gather points of higher probability [13].

Assuming that the number of infected individuals follows a well behaved concave function, we fitted a quadratic polynomial that represents this population at each week, according to Eq (29).

Thus, our prior for the compartment I (29) was governed by a second degree polynomial whose coefficients a (23) and b (27) are decentralized variables defined by normally distributed central values with mean 0 and standard deviation 10000 a_μ (20) and b_μ (24), plus a

perturbation characterized by $a_o \cdot a_\sigma$ and $b_o \cdot b_\sigma$ respectively and a coefficient c (28) normally distributed with mean equal to the $pim \cdot N$ and standard deviation 10000.

The constant ($pim = 58/16191$) is the percentage of infected at time ($t = 0$), calculated as the ratio between the number of infected at time ($t = 0$) on Moorea Island [7] and the total population of Moorea Island.

The offsets a_o (22) and b_o (26) are multidimensional standard normally distributed variables that scale the random variables a_σ (21) and b_σ (25), Half-Cauchy distributed variables that represents the standard deviation of a_μ and b_μ respectively. The Half Cauchy is a typical choice of prior for standard deviation variables to keep the inference sensitive to weakly-informative prior distribution [14].

Therefore the size of the population in the infectious class on the t -th week was summarized in the quadratic function (29), a decentralized procedure to make easier to the MCMC algorithm scan areas in the space with high curvature, as described by [13].

$$a_\mu = \mathcal{N}(0, 10000) \quad (20)$$

$$a_\sigma = \text{Cauchy}(0, 5) \quad (a_\sigma \geq 0) \quad (21)$$

$$a_o^t = \mathcal{N}(0, 1) \quad (22)$$

$$a = a_\mu + a_o * a_\sigma \quad (23)$$

$$b_\mu = \mathcal{N}(0, 10000) \quad (24)$$

$$b_\sigma = \text{Cauchy}(0, 5) \quad (b_\sigma \geq 0) \quad (25)$$

$$b_o^t = \mathcal{N}(0, 1) \quad (26)$$

$$b = b_\mu + b_o * b_\sigma \quad (27)$$

$$c = \mathcal{N}(pmi \cdot N, 1000) \quad (28)$$

$$I_t = a_t \cdot t^2 + b_t \cdot t + c \quad (29)$$

As the time interval (Δt_i) between points in our time series was constant and equal to a week, we can set (30) equal to 1.0 and carry out our study using finite differences.

$$\Delta t_i = 1, \quad (0 \leq i \leq n) \quad (30)$$

The approximation (31), from the equalities (3) and (30), made it evident that the t -th inflow to the R compartment is computed as the population in the I compartment on the t -th week divided by γ^{-1} .

$$\Delta R_w \approx (\gamma \cdot I_t) = \frac{I_t}{\gamma^{-1}} \quad (31)$$

Also, we know that once in the R -compartment, there is no outflow. So the population in the w -th week (R_w) is the population in the $(w - 1)$ -th week plus the input stream ($R_w = R_{w-1} + \Delta R_w$). In other words, R is the cumulative sum of the input stream in each period (week), given by the Eq (32) (assuming $R_0 = 0$).

Considering that the population in Rio de Janeiro (N) was assumed constant, the population S in the w -th week (S_w) was determined by the Eq (33).

Therefore, the incidence function in the study interval was given deterministically by Eq (34).

$$R_w = \sum_{i=1}^w \Delta R_i \quad (32)$$

$$S_w = (N - R_w - I_w) \quad (33)$$

$$\Delta S^w = \frac{\mathcal{R}_t \cdot S_w \cdot I_w}{\gamma^{-1} \cdot N} \quad (34)$$

The Table (1) summarizes the prior distributions used for females in fertile age model.

For the Male, Female and Total models, we have set the parameter values related to pregnancy (pp and mz) to the posterior means obtained in the Female in Fertile Age model.

Results

Epidemiological parameter estimates

The Fig 4 shows the posterior distributions of the probability of observing (reporting) a case of zika for each of the subgroups (Male—(a), Female—(b), Female in Fertile Age—(c) and Total Population—(d)) during 2016 in Rio de Janeiro. The probability of observation for men was noticeably lower than for women.

Fig 5 summarizes the posterior distributions of the infectious period for each of the subgroups (Male—(a), Female—(b), Female in Fertile Age—(c) and Total Population—(d)). As expected, the results indicated that, on average, the male and female infectious periods were the same. Also, the results indicated a small displacement to the right of what was established a priori, due to the mosquito life span and the infectious period by sexual transmission (implicitly attached in our model to the vector transmission process).

Our results indicated that Zika had a 42.6% (95%CI: 29.1–56.6%) participation in reduction of fertility in 2016 as shown in Fig 6.

We estimated the total incidence for each of the subgroups (Male—Fig 7, Female—Fig 8, Female in Fertile Age—Fig 9 and Total Population—Fig 10) reported and unreported in the analyzed period.

Table 1. Prior distributions of the parameters used in the model.

Parameter	Definition	Prior	References
po	Probability of report (observance)	Beta(17.4, 134.2)	[7]
mz	Percentage of fertility decline caused by Zika	Beta(12.0, 12.0)	[2]
γ^{-1}	Infectious Period	Gamma(14.0, 2.0)	[7]
\mathcal{R}_t	Reproductive Number	Gamma(15.4, 3.2)	[7]
pp	Probability of pregnancy	Beta(100.1, 1579.7)	
a_μ, b_μ	Mean of the parameters a and b	$\mathcal{N}(0, 10000)$	
a_σ, b_σ	Standard deviation of the parameters a and b	$Cauchy(0, 5)$	
a'_o, b'_o	Offset of the parameters a and b	$\mathcal{N}(0, 1)$	
c	Intercept of the Infectious compartment	$\mathcal{N}(pmi \cdot N, 10000)$	

<https://doi.org/10.1371/journal.pone.0205001.t001>

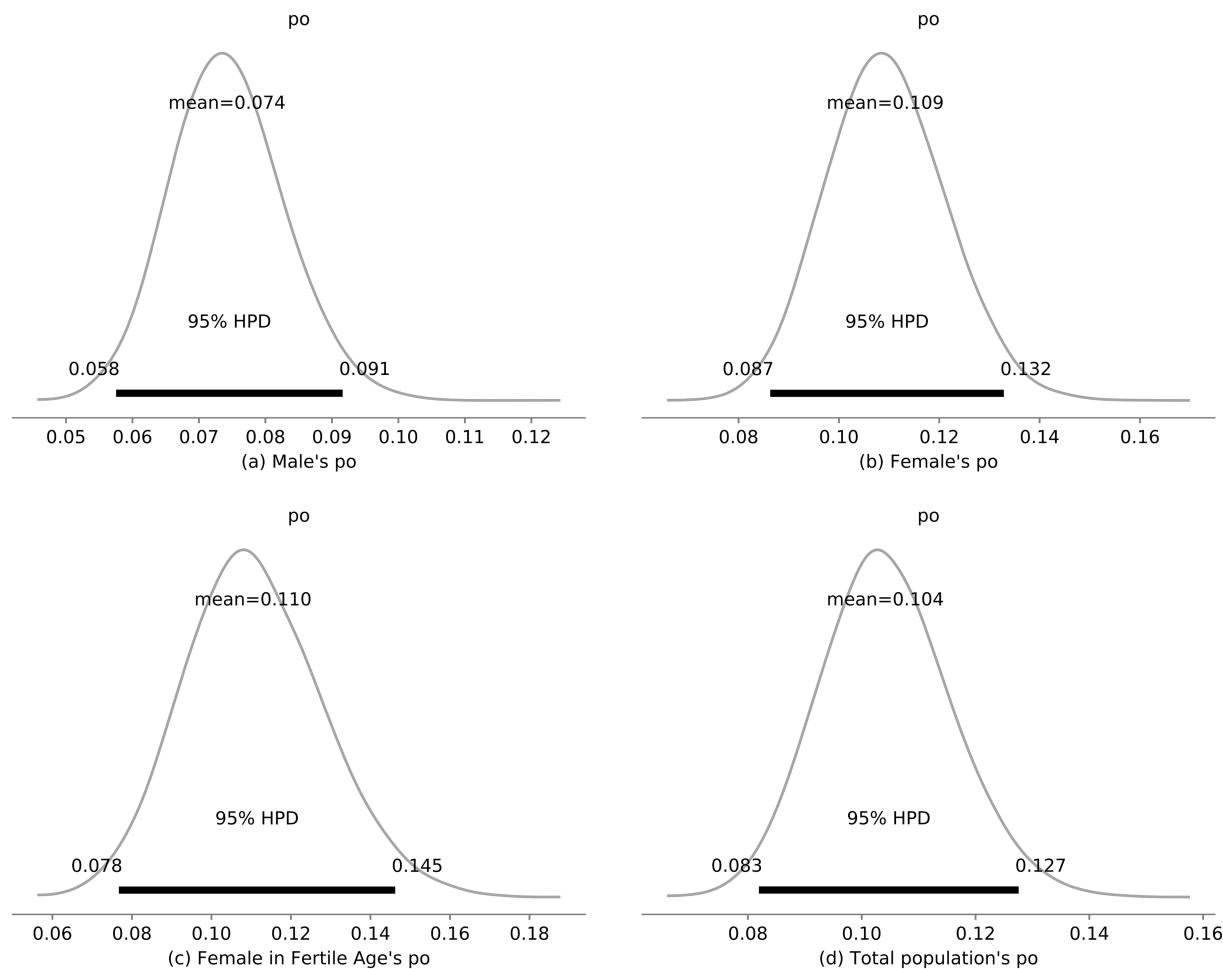


Fig 4. Probability of observation. Posterior probability distribution of the probability of observation in each sub-population model. In particular, the mean and highest 95% posterior density interval of each parameter.

<https://doi.org/10.1371/journal.pone.0205001.g004>

On Fig 11 we can see the distribution of the basic reproductive number \mathcal{R}_0 , for the Zika epidemic in total population of Rio de Janeiro city in 2016.

Discussion and conclusion

In this study, a new approach was developed for inferring the parameters that govern the spread of the Zika virus within Rio de Janeiro in 2016 and for estimating its real reach.

In our study, we found that only 10.4% (95% CI: 8.3–12.7%) of the total infections were reported. This result agrees with the estimated fraction of the population reported in other regions [7]. Also we verified that from the drop in the birth rate in 2016 compared to 2015, 42.6% (95% CI: 29.1–56.6%) was caused by Zika virus, and a really significant contribution caused by reasons like miscarriage and avoided pregnancies.

Our results show that a person remains infectious for 9.599 (95% CI: 7.673–11.701) days, which agrees with previous studies [7, 15], despite the fact that we did not include vector transmission in the model.

Knowing the full size of an epidemic is very relevant to understanding the full impact of a disease both in terms of morbidity and also in consequent economic losses. Particularly for

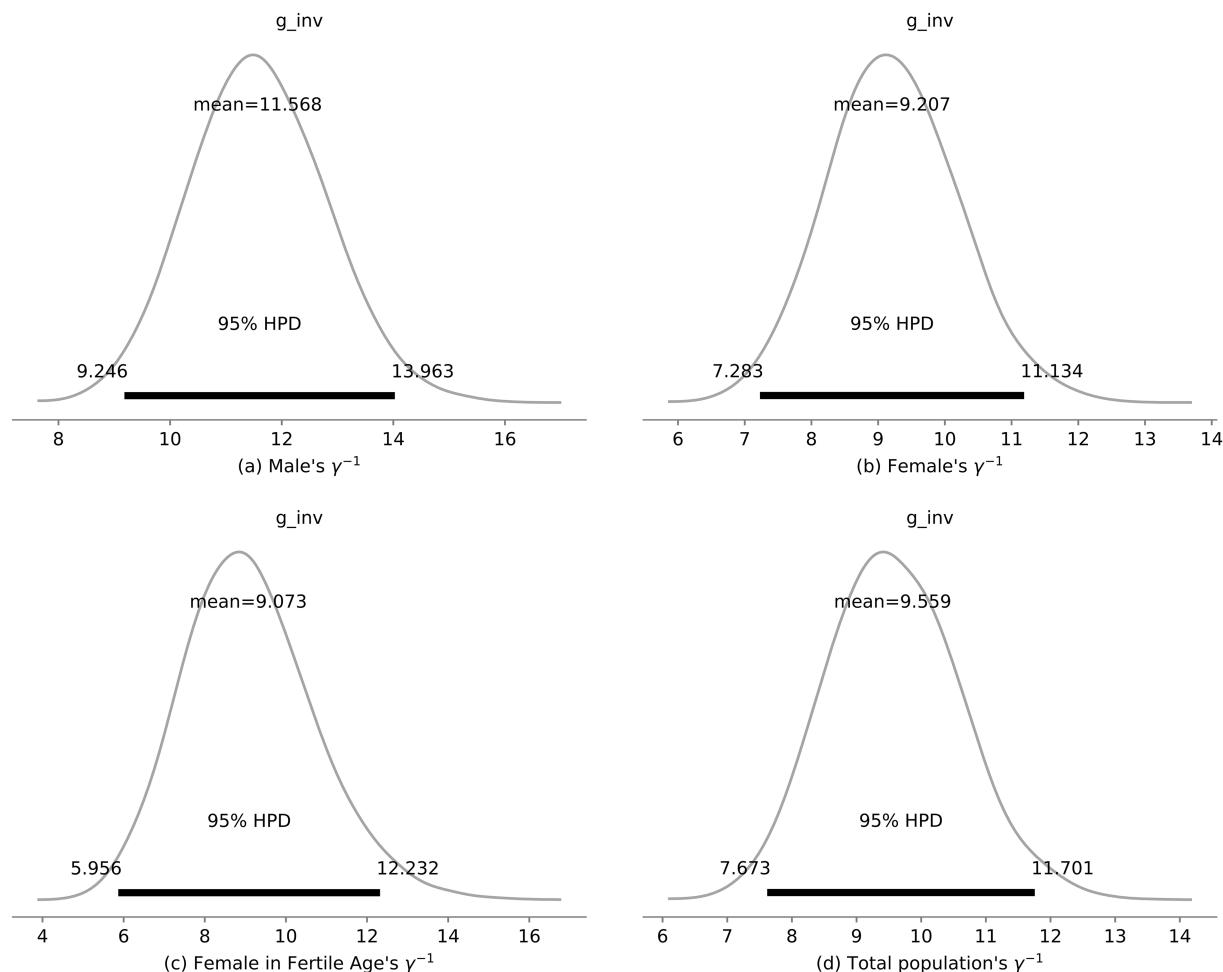


Fig 5. Infective period (γ^{-1}). Posterior probability distribution of the γ^{-1} in each sub-population model. In particular, the mean and highest 95% posterior density interval of each parameter.

<https://doi.org/10.1371/journal.pone.0205001.g005>

Zika, the asymmetry in its dissemination caused by sexual transmission from men to women of fertile age poses an additional challenge since we cannot assume the same transmissibilities for both sexes. Different perceptions of risk for men and women lead to different underreporting rates by gender for the epidemic as a whole. The impact of Zika on the early miscarriage rate turned the live-birth record into a key source of information allowing us to estimate the under-reporting rates. The fact that the birth record is complete made it possible for us to properly measure the epidemic.

Several favorable factors made it possible for us to study this phenomenon in Rio de Janeiro: detailed data at the level of individual cases, a large epidemic in a fully susceptible population and very stable fertility rates due to the large population.

The use of a SIR transmission model as a constraint for the inference on the fully realized dynamics without requiring the simulation of the entire model dynamics at each MCMC step, is an original approach, which greatly reduces the computational cost of the entire inferential process. We believe that using mechanistic transmission models as a basis for inferring latent parameters in the context of communicable disease dynamics, is a key to

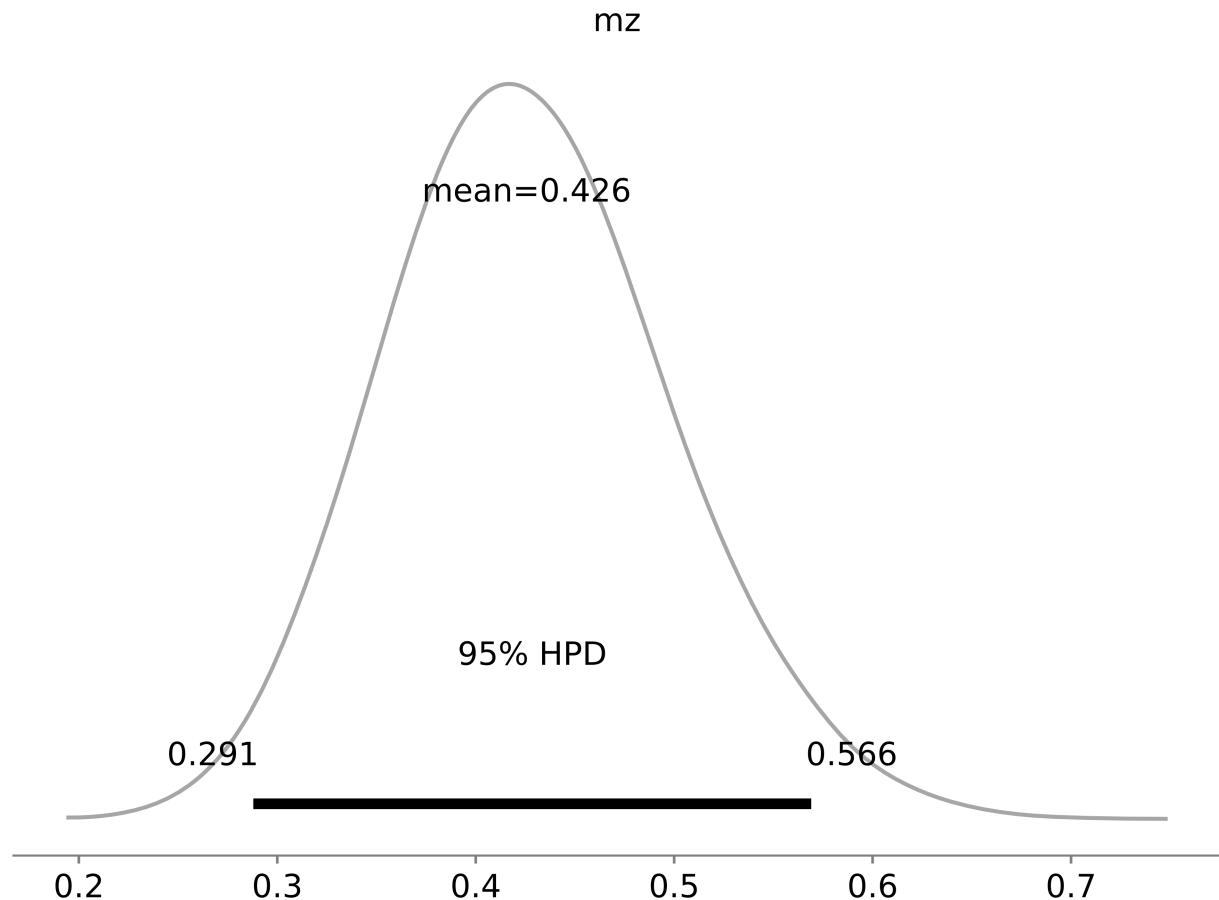


Fig 6. Estimated value of the quota on fertility reduction in 2016 due by Zika (mz).

<https://doi.org/10.1371/journal.pone.0205001.g006>

ensuring a causal interpretation of the results because these causal mechanisms are explicitly stated.

Although estimating \mathcal{R}_0 was not the main focus of this study, the values obtained (2.53% (95% CI: 1.59–3.45%)) are consistent with previous estimates [8, 16, 17].

Thus, although we applied a simpler epidemiological model, the statistical model was efficient enough to obtain results similar to those obtained in other studies with more detailed epidemiological models.

Knowing the full size of the Zika epidemic in Rio in 2016, is very important if we are to fully understand the risks of having a second wave of Zika in the city in the near future. In 2017, there was no outbreak of Zika in Rio, despite the continued low-intensity transmission [18].

Although our model presents general results similar to those of other studies, it does not provide interpretations of the mosquito's role in the vector transmission process, nor does it give the magnitude of the sexual contribution in the transmissive process. These are limitations of our work and on the understanding of the virus spreading dynamics.

In this paper we sought to provide answers to the very important issue of the true size of the epidemic. The next natural step would be to use the same dataset to generate estimates for the parameters of a full vector and sexual transmission model.

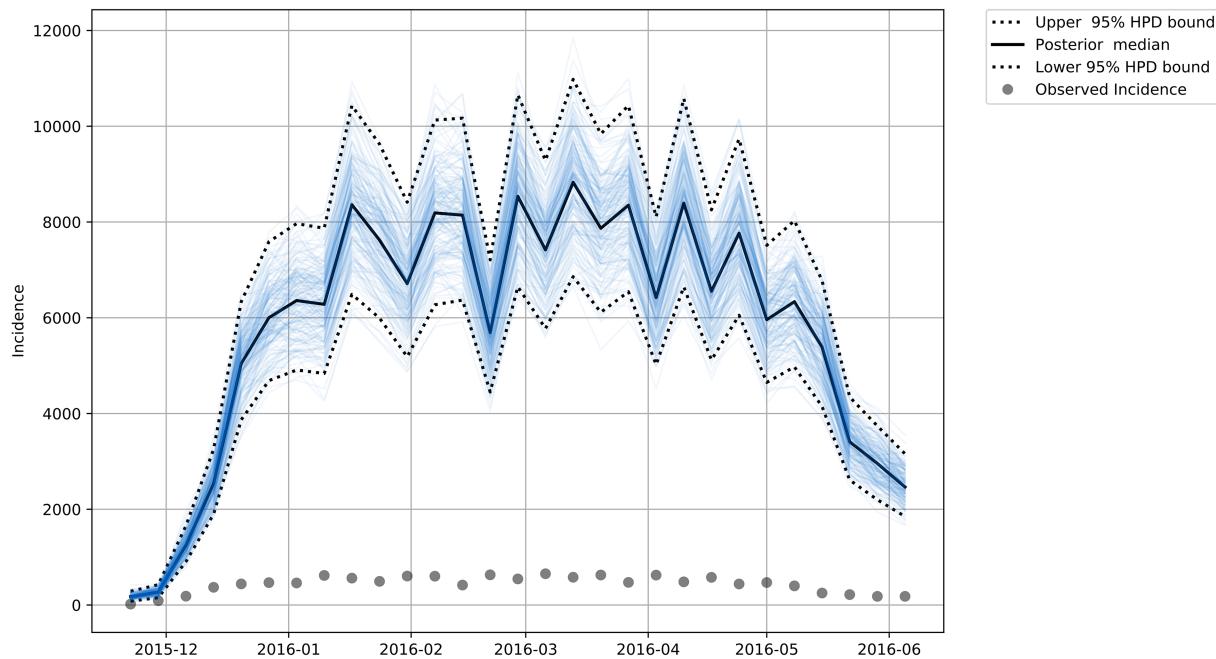


Fig 7. Estimated Total Zika male incidence. The dashed lines bounds the 95% credibility range and the continuous darker curve represents the median of the total male incidence. The dots are the observed incidences in our data.

<https://doi.org/10.1371/journal.pone.0205001.g007>

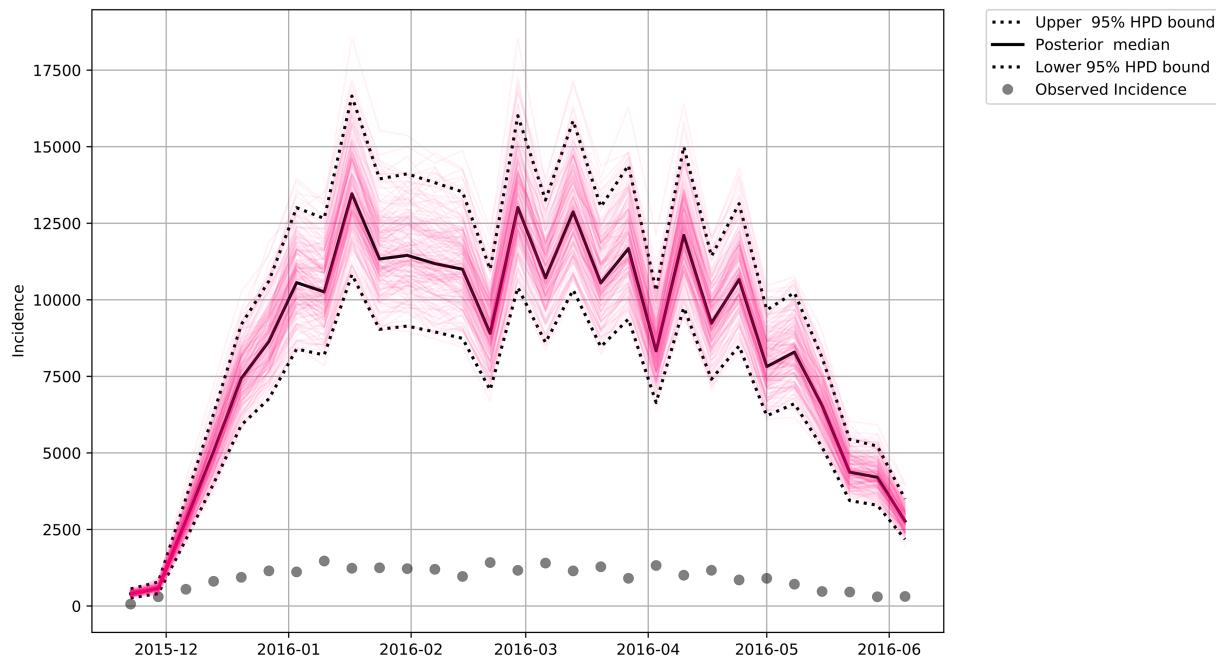


Fig 8. Estimated Total Zika female incidence. The dashed lines bounds the 95% credibility range and the continuous darker curve represents the median of the total female incidence. The dots are the observed incidences in our data.

<https://doi.org/10.1371/journal.pone.0205001.g008>

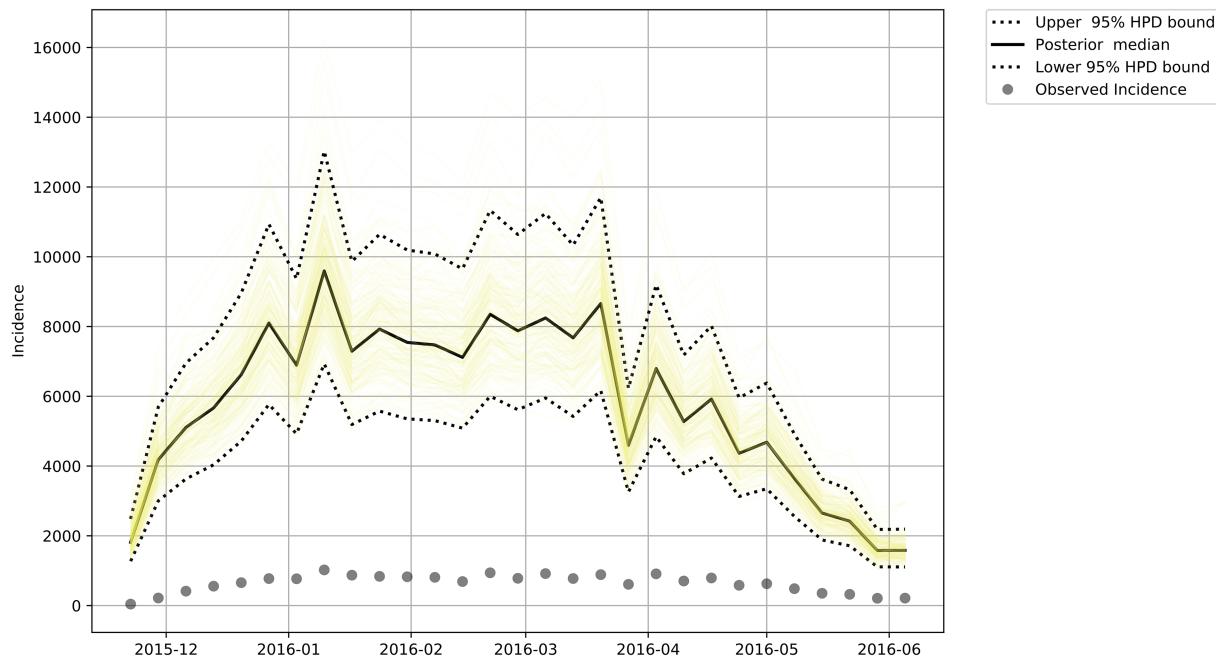


Fig 9. Estimated Total Zika Female in Fertile Age Incidence. The dashed lines bounds the 95% credibility range and the continuous darker curve represents the median of the total female in fertile age incidence. The dots are the observed incidences in our data.

<https://doi.org/10.1371/journal.pone.0205001.g009>

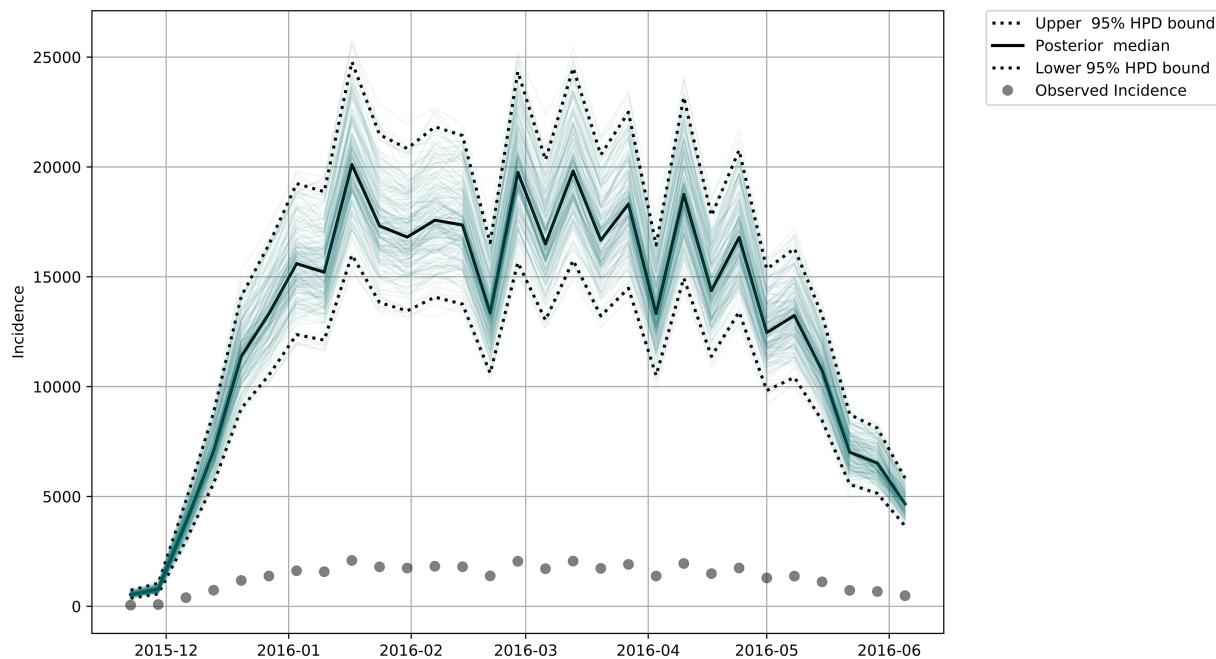


Fig 10. Estimated Total Incidence. The dashed lines bounds the 95% credibility range and the continuous darker curve represents the median of the total incidence. The dots are the observed incidences in our data.

<https://doi.org/10.1371/journal.pone.0205001.g010>

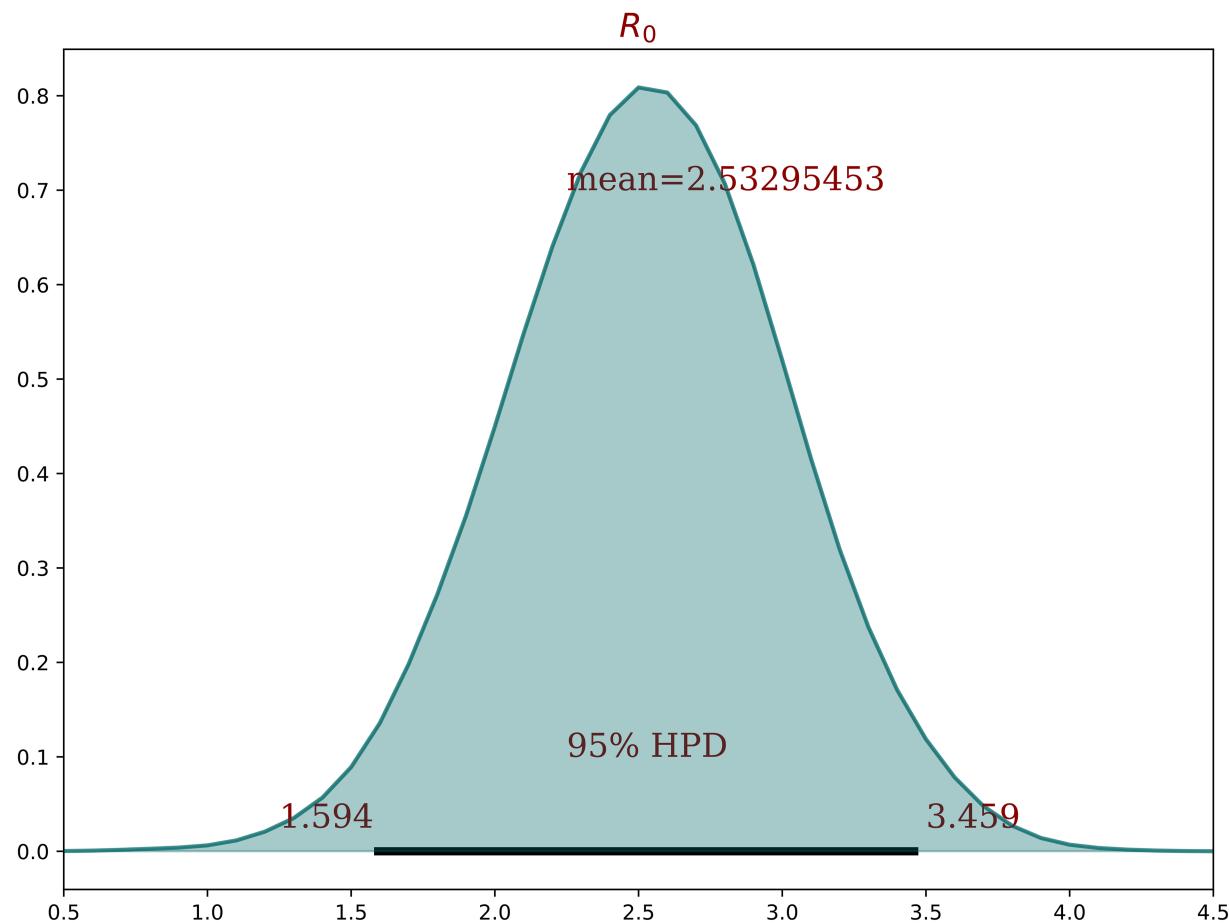


Fig 11. Basic reproductive number (R_0). Posterior distribution of the basic reproductive number of the Zika epidemic in Rio de Janeiro in 2016.

<https://doi.org/10.1371/journal.pone.0205001.g011>

Acknowledgments

The authors are grateful to the Infodengue project for making the datasets available. The authors also acknowledge the donation of TESLA GPU card from NVIDIA, without which the computations would have taken much longer.

Author Contributions

Data curation: Flávio Codeço Coelho.

Formal analysis: Marcio M. Bastos.

Funding acquisition: Flávio Codeço Coelho.

Methodology: Flávio Codeço Coelho.

Resources: Flávio Codeço Coelho.

Supervision: Flávio Codeço Coelho.

Visualization: Marcio M. Bastos.

Writing – original draft: Marcio M. Bastos.

Writing – review & editing: Flávio Codeço Coelho.

References

1. Coelho FC, Durovni B, Saraceni V, Lemos C, Codeco CT, Camargo S, et al. Higher incidence of Zika in adult women than adult men in Rio de Janeiro suggests a significant contribution of sexual transmission from men to women. *International Journal of Infectious Diseases*. 2016; 51:128–132. <https://doi.org/10.1016/j.ijid.2016.08.023> PMID: 27664930
2. Coelho FC, Armstrong M, Saraceni V, Lemos C. Can Zika account for the Missing Babies? *Frontiers in public health*. 2017; 5:317. <https://doi.org/10.3389/fpubh.2017.00317> PMID: 29238705
3. Zinszer K, Morrison K, Brownstein JS, Marinho F, Santos AF, Nsoesie EO. Reconstruction of Zika virus introduction in Brazil. *Emerging infectious diseases*. 2017; 23(1):91. <https://doi.org/10.3201/eid2301.161274> PMID: 27618573
4. Bogoch II, Brady OJ, Kraemer MU, German M, Creatore MI, Kulkarni MA, et al. Anticipating the international spread of Zika virus from Brazil. *The Lancet*. 2016; 387(10016):335–336. [https://doi.org/10.1016/S0140-6736\(16\)00080-5](https://doi.org/10.1016/S0140-6736(16)00080-5)
5. Heesterbeek JAP. A brief history of R₀ and a recipe for its calculation. *Acta biotheoretica*. 2002; 50(3):189–204. <https://doi.org/10.1023/A:1016599411804> PMID: 12211331
6. Hethcote HW. The mathematics of infectious diseases. *SIAM review*. 2000; 42(4):599–653. <https://doi.org/10.1137/S0036144500371907>
7. Kucharski AJ, Funk S, Eggo RM, Mallet HP, Edmunds WJ, Nilles EJ. Transmission dynamics of Zika virus in island populations: a modelling analysis of the 2013–14 French Polynesia outbreak. *PLoS neglected tropical diseases*. 2016; 10(5):e0004726. <https://doi.org/10.1371/journal.pntd.0004726> PMID: 27186984
8. Towers S, Brauer F, Castillo-Chavez C, Falconar AK, Mubayi A, Romero-Vivas CM. Estimate of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission. *Epidemics*. 2016; 17:50–55. <https://doi.org/10.1016/j.epidem.2016.10.003> PMID: 27846442
9. Coelho FC, De Carvalho LM. Estimating the attack ratio of dengue epidemics under time-varying force of infection using aggregated notification data. *Scientific reports*. 2015; 5:18455. <https://doi.org/10.1038/srep18455> PMID: 26675824
10. Paixão ES, Barreto F, da Glória Teixeira M, da Conceição N Costa M, Rodrigues LC. History, epidemiology, and clinical manifestations of Zika: a systematic review. *American journal of public health*. 2016; 106(4):606–612. <https://doi.org/10.2105/AJPH.2016.303112> PMID: 26959260
11. D'Ortenzio E, Matheron S, de Lamballerie X, Hubert B, Piorkowski G, Maquart M, et al. Evidence of sexual transmission of Zika virus. *New England Journal of Medicine*. 2016; 374(22):2195–2198. <https://doi.org/10.1056/NEJMc1604449> PMID: 27074370
12. Hoffman MD, Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*. 2014; 15(1):1593–1623.
13. Betancourt M, Girolami M. Hamiltonian Monte Carlo for hierarchical models. *Current trends in Bayesian methodology with applications*. 2015; 79:30.
14. Gelman A, et al. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*. 2006; 1(3):515–534. <https://doi.org/10.1214/06-BA117A>
15. Bastos L, Villela DA, Carvalho LM, Cruz OG, Gomes MF, Durovni B, et al. Zika in Rio de Janeiro: Assessment of basic reproduction number and comparison with dengue outbreaks. *BioRxiv*. 2016; p. 055475.
16. Nishiura H, Mizumoto K, Villamil-Gómez WE, Rodríguez-Morales AJ. Preliminary estimation of the basic reproduction number of Zika virus infection during Colombia epidemic, 2015–2016. *Travel medicine and infectious disease*. 2016; 14(3):274–276. <https://doi.org/10.1016/j.tmaid.2016.03.016> PMID: 27060613
17. Villela D, Bastos L, De Carvalho L, Cruz OG, Gomes M, Durovni B, et al. Zika in Rio de Janeiro: Assessment of basic reproduction number and comparison with dengue outbreaks. *Epidemiology & Infection*. 2017; 145(8):1649–1657. <https://doi.org/10.1017/S0950268817000358>
18. Coimbra BM, Coelho FC, Armstrong M, Saraceni V, Lemos C. Zika: An ongoing threat to women and infants. *BioRxiv*. 2017; p. 220640.