

Projecting the end of the Zika virus epidemic in Latin America: a modelling analysis

Kathleen M O'Reilly^{1,2}, Rachel Lowe^{2,3,4}, W John Edmunds^{2,3}, Philippe Mayaud⁵, Adam Kucharski^{2,3}, Rosalind M Eggo^{2,3}, Sebastian Funk^{2,3}, Deepit Bhatia⁶, Kamran Khan⁶, Moritz U Kraemer^{7,8}, Annelies Wilder-Smith¹, Laura C Rodrigues³, Patricia Brasil⁹, Eduardo Massad¹⁰, Thomas Jaenisch¹¹, Simon Cauchemez^{12,13,14}, Oliver J Brady^{2,3*}, Laith Yakob^{1,2*}

¹ Department of Disease Control, London School of Hygiene & Tropical Medicine, London, United Kingdom.

² Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom.

³ Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom.

⁴ Barcelona Institute for Global Health (ISGLOBAL), Barcelona, Spain

⁵ Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, United Kingdom.

⁶ Division of Infectious Diseases, University of Toronto, Toronto, Ontario, Canada; and Centre for Research on Inner City Health, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Ontario, Canada.

⁷ Harvard Medical School, Harvard University, Boston, MA, USA; Boston Children's Hospital, Boston, MA, USA.

⁸ Department of Zoology, University of Oxford, Oxford, UK.

⁹ Instituto Nacional de Infectologia Evandro Chagas / Fiocruz, Rio de Janeiro, Brazil.

¹⁰ School of Applied Mathematics, Fundacao Getulio Vargas, Rio de Janeiro, Brazil..

¹¹ Department for Infectious Diseases and Parasitology, Department for Infectious Diseases, University of Heidelberg, Heidelberg, Germany.

¹² Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Paris, France.

¹³ Centre National de la Recherche Scientifique, URA3012, Paris, France

¹⁴ Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris, France

* Contributed equally

Abstract

Background Zika virus (ZIKV) emerged in Latin America & the Caribbean (LAC) region in 2013, and has had serious implications for population health in the region. In 2016, the World Health Organization declared the ZIKV outbreak a Public Health Emergency of International Concern following a cluster of associated neurological disorders and neonatal malformations. In 2017, Zika cases declined, but future incidence in LAC remains uncertain due to gaps in our understanding, considerable variation in surveillance and a lack of a comprehensive collation of data from affected countries.

Methods Our analysis combines information on confirmed and suspected Zika cases across LAC countries and a spatio-temporal dynamic transmission model for ZIKV infection to determine key transmission parameters and projected incidence in 91 major cities within 35 countries. Seasonality was determined by spatio-temporal estimates of *Aedes aegypti* vector capacity. We used country and state-level data from 2015 to mid-2017 to infer key model parameters, country-specific disease reporting rates, and the 2018 projected incidence. A 10-fold cross-validation approach was used to validate parameter estimates to out-of-sample epidemic trajectories.

Results There was limited transmission in 2015, but in 2016 and 2017 there was sufficient opportunity for wide-spread ZIKV transmission in most cities, resulting in the depletion of susceptible individuals. We predict that the highest number of cases in 2018 within some Brazilian States (Sao Paulo and Rio de Janeiro), Colombia and French Guiana, but the estimated number of cases were no more than a few hundred. Model estimates of the timing of the peak in incidence were correlated ($p<0.05$) with the reported peak in incidence. The reporting rate varied across countries, with lower reporting rates for those with only confirmed cases compared to those who reported both confirmed and suspected cases.

Conclusions The findings suggest that the ZIKV epidemic is by and large over, with incidence projected to be low in most cities in LAC in 2018. Local low levels of transmission are probable but the estimated rate of infection suggests that most cities have a population with high levels of herd immunity.

Keywords: Zika virus (ZIKV), epidemic, mathematical modelling, Latin America and the Caribbean (LAC), connectivity

Introduction

Starting as early as 2013,^{1,2} the Zika virus (ZIKV) invaded northeast Brazil and began to spread in the Latin America & Caribbean (LAC) region. The subsequent discovery of a cluster of Guillain-Barré syndrome cases and the emergence of severe birth defects led the World Health Organisation (WHO) to declare the outbreak a Public Health Emergency of International Concern in early 2016. The virus has since spread to 49 countries and territories across the Americas where autochthonous transmission has been confirmed.³

However, 2017 saw a marked decline in reported Zika cases and its severe disease manifestations.⁴ This decline has been widely attributed to the build-up of immunity against ZIKV in the wider human population,⁵ although it remains unknown how many people have been infected. To date, there has been limited use of population-based surveys to determine the circulation and seroprevalence of ZIKV in LAC, owing to challenges in interpretation of serological tests that cross-react with other flaviviruses (e.g. dengue).^{6,7} In addition to the reduction in Zika cases there has also been a marked reduction in incidence of reported dengue and chikungunya cases in Brazil, meaning that the role of climatic, and other factors affecting mosquito density, or cross-immunity between arboviruses cannot be ruled out.

While the decline in ZIKV incidence is undoubtedly a positive development, it exposes clear gaps in our understanding of its natural history and epidemiology, which limit our ability to plan for, detect, and respond to future epidemics. The short duration of the epidemic and the long lead time needed to investigate comparatively rare congenital impacts has meant maternal cohort studies, in particular, may be statistically underpowered to assess relative risk and factors associated with ZIKV-related adverse infant outcomes.⁸ The evaluation of the safety and efficacy of ZIKV vaccine candidates⁹ are now also faced with an increasingly scarce number of sites with sufficient ZIKV incidence.^{10,11}

There is an urgent need to predict which areas in LAC remain at risk of transmission in the near future, and estimate the trajectory of the epidemic. Projections can help public health policymakers plan surveillance and control activities, particularly in areas where disease persists. They can also be used by researchers, especially those in vaccine and drug development to update sample size calculations for ongoing studies to reflect predicted incidence within the time-window of planned trials. The findings identified from a continental analysis of ZIKV in LAC may be useful should ZIKV emerge in other settings, such as quantifying the spatial patterns of spread and impact of seasonality on incidence.

Several mathematical and computational modelling approaches have been developed to forecast continental-level ZIKV transmission^{5,11–14}. The focus has largely been on estimating which areas are likely to experience epidemic growth. It is apparent from the incidence in 2017 that many countries no longer reported an increasing incidence of cases. Due to either data unavailability or inaccuracies in the reported number of Zika cases in each country at the time of analysis, such approaches have either not used incidence data at all,^{15–17} fit models to data on other arboviruses¹⁴ or used selected Zika-related incidence data from particular countries^{5,12,13,18–21} to calibrate their models. Additionally, only a small number of studies have validated their model findings, either through comparison to

serological surveys or comparing model outputs to incidence data not used within model fitting.^{13,19-21} Considerably more data are now available across LAC and spanning multiple arboviral transmission seasons. This provides a valuable opportunity to examine the nature of ZIKV transmission and the importance of connectivity and seasonality in assessing ZIKV persistence in specific locations throughout LAC.

In this article, we apply a dynamic spatial model of ZIKV transmission in 91 major cities across LAC and fit the model to the latest data from 35 countries in LAC. We test several models to account for human mobility to better understand the impact of human movements on the emergence of ZIKV. The model was validated using a 10-fold cross-validation comparison to the data. We use the fitted model to quantify the expected number of cases likely to be observed in 2018 and identify cities likely to remain at greatest risk.

Methods

Zika case data from Latin America & the Caribbean

The weekly number of confirmed and suspected Zika cases within each country is reported to the Pan American Health Organization. This analysis makes use of the weekly incidence of Zika cases in 35 countries, from January 2015 to August 2017 (see Supporting Information S1). State-level ZIKV incidence data was available for Brazil and Mexico.²² Confirmed cases are typically identified through a positive real-time reverse polymerase chain reaction blood test using ZIKV-specific RNA primers. Suspected cases are based on the presence of pruritic (itchy) maculopapular rash together with two or more symptoms such as fever, or polyarthralgia (multiple joint pains), or periarticular oedema (joint swelling), or conjunctival hyperaemia (eye blood vessel dilation) without secretion and itch.^{23,24} Confirmed and suspected cases were included in this analysis because ZIKV detection may have low sensitivity due to a narrow window of viraemia and many samples, particularly from the earlier phase of the epidemic remain untested due to laboratories being overloaded during the epidemic.²⁴ Inclusion of suspected cases in the analysis may reduce specificity due to the non-specific clinical manifestations of ZIKV and similar circulating arboviruses, including dengue. The reporting of ZIKV cases will vary considerably between settings and is thought to depend on the arbovirus surveillance system already in place, additional surveillance specifically established for ZIKV and other viruses and the likelihood of an individual self-reporting with symptoms consistent with ZIKV infection.

A mathematical model of ZIKV infection

A deterministic meta-population model was used for ZIKV transmission between major cities in the LAC region. Cities with a population larger than 750,000 and large Caribbean islands were included in the model. In total we considered 91 cities. We extracted population sizes using the UN estimates from 2015.²⁵ Migration between cities was modelled assuming several scenarios: i) a gravity model with no exponential terms; ii) a gravity model with estimated exponential terms; iii) a radiation model; iv) a data-driven approach based on flight data; and v) a model of local radiation and flight movements. Gravity models assume that movement between cities is highest when located near each other and when both cities are large. Radiation models assume that movement between cities are affected by the size of the population in a circle between the cities (see Supporting Information S2 for further information).

Within each city, individuals were classified by their infection status: susceptible, pre-infectious, infectious or recovered from ZIKV infection (Figure 1). Upon infection, individuals were assumed to be pre-infectious for an average of 5 days and then infectious for a subsequent 20 days.^{26,27} Immunity was assumed to be life-long and no cross-protection against other flaviviruses was assumed. The main vectors for ZIKV in LAC are thought to be *Aedes aegypti*, whilst *Aedes albopictus* and other species were thought to play a minor role in transmission.²⁸ The seasonality and scale of ZIKV transmission was assumed to be specific to each city and dependent across cities, using a vector capacity modelling approach to estimate transmission intensity of the vectors and an environmental niche model to estimate vector relative abundance.^{29–31} In this approach, we model the probability that ZIKV may transmit for each day of the year, and feed this time-varying probability into the

mathematical model. We estimate the time-varying reproduction number ($R_0(t)$), defined the average number of secondary infections that result from one infected person within a totally susceptible population, that value varies in time due to the seasonality in vector capacity within each city. The seasonality curves are summarised using the average number of days per year where $R_0(t)$ was greater than 1, and the mean value of $R_0(t)$.

Owing to the difficulties in ZIKV disease surveillance,²³ the weekly incidence of reported cases was unlikely to reflect the true incidence in each setting and we did not fit the model to weekly incidence data. We therefore used summary statistics in the model fitting procedure, focussing on the timing of the peak in incidence and whether the annual incidence was above 1 case per 100,000 in each country. The timing of the peak in outbreaks has been previously shown to be a useful summary statistic for epidemic dynamics,^{32,33} and preliminary analysis illustrated that annual incidence had a good discriminatory power for the estimating parameters of the model. Although surveillance quality varies between settings the timing of the reported peak within countries is less sensitive to systematic error. A sensitivity analysis confirmed that only a small number of observations were susceptible to large changes in surveillance prior to April 2016 and after January 2017, making the reported timing of the peak robust to changes in surveillance (Supporting Information S3).

The model estimate of new infections within each city was aggregated to the country or state level (for Brazil and Mexico) and scaled to ZIKV cases, enabling comparisons with the available data. The maximal value of $R_0(t)$ and the best-fitting migration model (including the maximal leaving rate from cities) were estimated in the model fitting procedure. Parameters were estimated using approximate Bayesian computation - sequential Monte Carlo (ABC-SMC) methods.³⁴ ABC methods use summary statistics to estimate model parameters from qualitative epidemic characteristics. The sequential procedure of ABC-SMC means that each model of human mobility could be treated as a parameter. The prior and posterior distributions of selecting each model was used to estimate Bayes Factors to determine the evidence in favour of one model over another. Multiple parameter sets with equivalent fit were produced during the model fitting, and were used to provide the mean and 95% credible intervals (CI) of parameter estimates, numbers infected between 2015-2017, timing of the peak in the epidemic and projections of the numbers of ZIKV cases in 2018. The distribution of the timing of the peak was compared to the data using Bayesian posterior checks. The values correspond to probability that the data take a value less than or equal to the cumulative distribution function of the model, and values between 0.01-0.99 can be interpreted as evidence that the data and model estimate come from the same distribution. For each country the time series of reported cases were compared to the normalised model incidence. We compare the total number of reported cases to the estimated median (and 95% CI) number of infections to estimate the country-specific probability of reporting a case per infection.

To validate the parameter estimates and model output a cross-validation approach was used. The data was split into ten randomly allocated groups by country, each group was sequentially excluded from the parameter estimation procedure and the peak timing of the out-of-sample parameter estimates were compared to the data. The 95% CI of the cross-validated estimates were compared to

the within-sample peak estimates. For the 2018 projections we reported the median number of cases, accounting for the estimated reporting rate and uncertainty in model output. The 95% prediction interval had a variance equal to the sum of the variance of the model prediction and the variance of the expected value assuming a Poisson distribution. Comparison of 2018 predictions to data were not possible as data from affected countries have not been made publicly available (as of 2 May 2018).

To date, there is no evidence to suggest sexual transmission contributes to population transmission,^{35,36} so this transmission route was not considered in the model. Due to current unexplained variability,³⁷ we do not project the expected numbers of neonatal malformations or neurological disorders, such as microcephaly, associated with ZIKV infection.

Results

A gravity model, which assumes migration scales with large populations that are closely located to one another, provided the best fit the data (Table 1). We identified substantial spatial heterogeneity in transmission (country summaries are provided in Table 2); the average estimated value of R_0 was 1.81 (95% CI 1.74-1.87) and the average number of days per year where $R_0(t)>1$ was 253 days (95% CI 250-256 days). The average number of days where $R_0(t)>1$ varied from 116 days days (Costa Rica) to almost year-round transmission (several cities within Brazil (Belem & Salvador) and Colombia (Medellin & Cali), and Aruba and Curacao Islands). The maximal estimate of $R_0(t)$ was often higher within cities that also reported a longer window of transmission with $R_0(t)>1$. However, several cities (including Boa Vista, Aracaju and Natal in Brazil) were estimated to have maximal $R_0(t)$ values above 2.5 with a relatively small window of transmission within the year.

Despite the emergence of the ZIKV epidemic in early 2015 in north-eastern Brazil, the incidence of cases remained relatively low in 2015 (Figure 2D and S5 for plots of Brazilian States). All countries that reported cases in 2015 (Brazil, Colombia, Guatemala, Honduras, Paraguay, Suriname, Cuba, El Salvador, Mexico, and Venezuela) continued to report cases in 2016 and 2017, except for Cuba. For most countries, the largest number of cases were reported in 2016. Belize, Colombia, French Guiana, Honduras, Suriname and several Caribbean islands reported more than 2 cases per 1,000 population in 2016. For 28 of the 35 countries in the analysis, the peak in reported disease incidence occurred in 2016. Five countries reported a peak in 2017 and Cuba reported a peak in July 2015 (Figure 2C).

The estimated incidence of ZIKV infections (median and 95% CI) were compared to the reported data to estimate the country specific reporting rate. The average probability of an infection being reported as a case was 3.9% (95% CI 2.3-8.1%) and this rate was lower within countries that only reported confirmed cases (4 countries) than those who reported both confirmed and suspected cases (22 countries) (Table 2). Costa Rica, French Guiana and the US Virgin Islands were estimated to have a reporting rate above 20%. A comparison of the time-series of reported cases was compared to the model estimates of incidence (Figure 3). For all countries an epidemic was likely to have begun by Dec 2015 to Mar 2016 (otherwise known as the first phase). The relative scale of the epidemic in the first phase compared to late 2016 (the second phase) varied by country. For many countries the epidemic was estimated to be larger during the first phase (such as Argentina, Bolivia, Ecuador, Paraguay). For simulations in Antigua and Barbuda, Mexico and Venezuela the epidemic during the second phase had a higher incidence than the first phase. A small number of countries were estimated to have experienced only one epidemic season: Belize, Honduras, El Salvador and most Caribbean Islands. The difference in the timing of the peak between data and model was measured using Bayesian posterior checks where there was a non-significant difference between the model and data for 11 countries, and

the distribution was over-dispersed (Figure 4A and 4B). There was a significant correlation ($p=0.035$) between the reported and estimated peak in the country epidemics (Figure 4C). The estimated peak in cross-validated simulations were correlated ($p<0.001$) with the model fit, although the 95% CI were wider (Figure 4D).

Projections for 2018 suggest a low incidence of Zika cases in most cities considered in the analysis (Figure 5, and Table 2). When accounting for the country-specific case reporting rate, the median number of cases was typically less than 20 in most settings. However, French Guiana was predicted to have between 148-1773 cases, owing to a larger pool of susceptible individuals than in other settings. Populated States within Brazil, such as Santa Carina and São Paulo were projected to have more than 5 cases, and cases were predicted to occur within Medellin (Colombia) and San Jose (Costa Rica). The majority of Caribbean countries were predicted to have no cases in 2018. For all cities the incidence of cases in 2018 will be lower than 2017. In Colombia, the projected time series of cases for specific cities illustrate a negligible incidence in 2018, but Medellin was expected to experience the end of the epidemic in 2018 (Figure 5C).

Discussion

The spread of ZIKV across the LAC region in 2015-2017 has resulted in considerable disease burden, particularly in the children of mothers infected during pregnancy. Both the reported incidence of cases and modelling results from this study suggest that the transmission of ZIKV had continued until herd immunity was reached, despite major efforts to limit its spread through vector control. Whilst the reported and projected reduction in ZIKV cases is undoubtedly good news for affected communities, it is only because substantial numbers of individuals have already been infected. Therefore, it remains vital to maintain surveillance for congenital and developmental abnormalities and provide long-term care for affected people and families.³⁸

The aim of this analysis was to assess if cities in LAC were likely to experience ZIKV cases in 2018, to support resource planning and trials. Our modelling results suggest a very low incidence in 2018. This analysis supports the findings of previous mathematical models of ZIKV.^{5,11,13,14} In addition, our study provides estimates of incidence and risk for specific cities, estimates of case reporting rates, incorporates parameter uncertainty, includes out-of-sample validation of the model estimates and uses more data than other modelling studies as we incorporate ZIKV case reports alongside ecological data to determine city-specific epidemic trajectories and seasonality curves.

We fitted the model to the timing of the peak in ZIKV cases and then compare the time series of expected cases to reported cases and found a good fit in many countries. We assumed that large cities both drive the spread of Zika and are responsible for the majority of cases. As *Ae. aegypti* is a largely urban dwelling mosquito and that arboviral diseases have been observed to be spread by movement of infected humans,^{39,40} this assumption is likely to be valid. However, while we predict the outbreak to largely be over in these large cities, smaller more remote cities and peri-urban areas may still have susceptible individuals and experience cases. Should additional sub-national data on the timing of the peak become available, the model fitting and projections can easily be updated. Case reporting rates indicate a lower rate within countries that report only confirmed cases, and the rates within Brazil, El Salvador, Martinique, Puerto Rico, and Suriname align well with other estimates measured using alternative methods.^{21,41,42}

Despite the short-comings in the available data, we present the most up-to-date and robust predictions of Zika incidence in 2018. As the projected incidence is consistently low across all model runs, this finding is quite robust to the variability accounted for in the model. Validation of these findings are necessary through multi-site population representative seroprevalence surveys across LAC to monitor seroconversion to ZIKV, such as in Netto et al..¹⁹ Reporting of cases within LAC has reduced markedly since the downgrading of ZIKV from a public health emergency of international concern to an ongoing public health challenge (in November 2018).⁴³ Consequently, it remains difficult to compare these projections to incidence data for 2018.

This research has highlighted that within LAC the spread of ZIKV was better represented by a gravity model than flight movements. This may seem surprising as flight data are cited as a source of emerging infections, such as ZIKV.⁴⁴ But cars and public transportation are used for most journeys; and the movement of people impacts the spatial spread of vector-borne diseases.^{38,45} Perhaps for highly transmissible infectious diseases movements facilitated by flights are sufficient for predicting introduction of a pathogen in a new population, but this analysis suggests that triggering of a ZIKV outbreak may require more frequent exposure than air travel. The migration patterns assumed within each model are quite different in LAC (see Supporting Information), suggesting that models which have not tested the relative fit of each and use one alone could be prone to errors in estimated spread of ZIKV. In comparison to mobility modelling in North America, Europe and Africa, the mobility patterns in Latin America are not well quantified and require further study.

Major questions on the epidemiology of ZIKV remain unanswered.⁷ Whilst the impact of sexual transmission on ZIKV emergence is likely to be minimal,^{35,46} it may increase the magnitude of an epidemic³⁶ and this would be difficult to test using the available surveillance data. There are large differences in the incidence of congenital Zika syndrome across LAC,³⁸ with an epicentre reported within northeast Brazil, that remain largely unexplained. In particular, the analysis here suggests increased incidence of Zika throughout Brazil in 2016, but the expected increase in congenital malformations within newborns were not observed.⁴⁷ This and other modelling studies suggest that ZIKV has been widespread, and the finding of geographically variable rates of congenital defects is discordant with the more consistent rates of ZIKV infection predicted by our model.

We have assumed that the time varying transmission rate of ZIKV is a function of environmental and vector suitability that has not been reduced by effective vector control. The impact of vector control has been largely unassessed, or where it has been assessed it has been found to be ineffective.^{48,49} Consequently, our findings are likely to be unaffected by the impact of vector control. Should effective wide-scale interventions be developed, the model can be used to assess the impact of proposed interventions. The mathematical model was deterministic in nature, and especially for projections may under-estimate the variability in the number of cases. Additionally, we do not include the impact of inter-annual variation in *Aedes aegypti* vector capacity, such as the 2015-2016 El Nino climate phenomenon, which has previously been shown to be positively associated with an increased incidence in 2016.¹⁸ Instead, we show that the peak incidence in 2016 was likely due to a low incidence of infection in 2015, that then resulted in optimal transmission in 2016, resulting in depletion of the susceptible population, thus limiting incidence in 2017 and 2018. If inter-annual variation in ZIKV transmission were incorporated into our model, it is likely that our incidence estimates for 2016 would increase, and the predicted incidence in subsequent years would further decrease.

Conclusions

ZIKV has spread widely across LAC, affecting all cities during 2015-2017 leading to high population immunity against further infection, thereby limiting capacity for sustained ZIKV transmission. The seasonality in ZIKV transmission affected the rate of infection, but due to high connectivity between cities this had little impact on the eventual depletion of susceptible populations. Looking forward, incidence is expected to be low in 2018. This provides optimistic information for affected communities, but limits our ability to use prospective studies to better characterise the epidemiology of ZIKV. The continental-wide analysis illustrates much commonality between settings, such as the relative annual incidence, and the connectivity across LAC, but questions remain regarding the interpretation of the varied data for ZIKV. Ultimately, representative seroprevalence surveys will be most useful to understanding past spread and future risk of ZIKV epidemics in LAC.

Abbreviations

CI - credible intervals

LAC – Latin America and the Caribbean

ZIKV – Zika virus

Declarations

Ethics approval

Institutional ethics approval was not sought because this is a retrospective study and the databases are anonymised and free of personally identifiable information.

Availability of data and material

The models developed here and data are available at <https://github.com/kath-o-reilly/Zika-LAC-Outbreaks>.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was partially supported by the European Union's Horizon 2020 Research and Innovation Programme under ZIKAlliance Grant Agreement no. 734548 and ZikaPLAN Grant Agreement No 734584. RL was funded by a Royal Society Dorothy Hodgkin Fellowship. AK was funded by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant Number 206250/Z/17/Z). OJB was funded by a Sir Henry Wellcome Fellowship funded by the Wellcome Trust (grant number 206471/Z/17/Z). SC was funded by the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases program (grant ANR-10-LABX-

62-IBEID), the Models of Infectious Disease Agent Study of the National Institute of General Medical Sciences, the AXA Research Fund.

Authors contributions

KO, RL, JE, PM, AK, RE, SC, OB and LY designed the study and KO, RL, DB, KK, OB and LY carried out the analysis. KO, RL, JE, PM, AK, RE, SF, DB, KK, MK, AW, LR, PB, EM, TJ, SC, OB and LY contributed to the data analysis and interpretation, writing of the report and approved it before submission.

Acknowledgements

We thank partners within the ZikaPlan and Zika Alliance consortia who have provided comments and insight on previous versions of this work. We also greatly acknowledge the availability of publicly accessible data on ZIKV cases and other inputs of the model.

References

- 1 Faria NR, Quick J, Claro IM, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 2017; **546**: 406–10.
- 2 Faria NR, Azevedo RDSD, S, Kraemer MUG, et al. Zika virus in the Americas: Early epidemiological and genetic findings. *Science* (352) 2016; **352**: 345–9.
- 3 WHO. The Zika Strategic Response Plan. 2016. <http://apps.who.int/iris/bitstream/handle/10665/246091/WHO-ZIKV-SRF-16.3-eng.pdf;jsessionid=2B3E1DFBBFCA23BB42E351E4941509D0?sequence=1>
- 4 WHO. Zika situation report - 10 March 2017. 2017. <http://www.who.int/emergencies/zika-virus/situation-report/10-march-2017/en/>
- 5 Ferguson NM, Cucunuba ZM, Dorigatti I, et al. Countering the Zika epidemic in Latin America. *Science* 2016; **353**: 353–4.
- 6 Lanciotti RS, Kosoy OL, Laven JJ, et al. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg Infect Dis* 2008; **14**: 1232–9.
- 7 Aliota MT, Bassit L, Bradrick SS, et al. Zika in the Americas, year 2: What have we learned? What gaps remain? A report from the Global Virus Network. *Antiviral Res* 2017; **144**: 223–46.
- 8 Butler D. Drop in cases of Zika threatens large-scale trials. *Nature* 2017; **545**: 396–7.
- 9 WHO. WHO vaccine pipeline tracker - ZIKV. 2017. http://www.who.int/immunization/research/vaccine_pipeline_tracker_spreadsheet/en/ (accessed Nov 19, 2017).
- 10 Durbin A, Wilder-Smith A. An update on Zika vaccine developments. *Expert Rev Vaccines* 2017; **16**: 781–7.
- 11 Asher J, Barker C, Chen G, et al. Preliminary modeling results for Zika virus transmission in 2017. *bioRxiv* 2017 <https://doi.org/10.1101/187591>.
- 12 Rodriguez-Barraquer I, Salje H, Lessler J, Cummings DA. Predicting intensities of Zika infection and microcephaly using transmission intensities of other arboviruses. *bioRxiv* 2016 <https://doi.org/10.1101/041095>.
- 13 Zhang Q, Sun K, Chinazzi M, et al. Spread of Zika virus in the Americas. *Proc Natl Acad Sci* 2017; **114**: E4334–43.
- 14 Colón-González FJ, Peres CA, Steiner São Bernardo C, Hunter PR, Lake IR. After the epidemic: Zika virus projections for Latin America and the Caribbean. *PLoS Negl Trop Dis* 2017; **11**: e0006007.
- 15 Bogoch II, Brady OJ, Kraemer MUG, et al. Anticipating the international spread of Zika virus from Brazil. *Lancet* 2016; **387**: 335–6.
- 16 Brady OJ, Godfray HCJ, Tatem AJ, et al. Vectorial capacity and vector control: Reconsidering sensitivity to parameters for malaria elimination. *Trans R Soc Trop Med Hyg* 2016; **110**: 107–17.
- 17 Perkins AT, Siraj AS, Ruktanonchai CW, Kraemer MUG, Tatem AJ. Model-based projections of Zika virus infections in childbearing women in the Americas. *Nat Microbiol* 2016; **1**: 16126.
- 18 Caminade C, Turner J, Metelmann S, et al. Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. *Proc Natl Acad Sci* 2017; **114**: 119–24.
- 19 Netto EM, Moreira-Soto A, Pedroso C, et al. High Zika Virus Seroprevalence in Salvador, Northeastern Brazil Limits the Potential for Further Outbreaks. *MBio* 2017; **8**: e01390-17.
- 20 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 Negl Trop Dis.* 2016 May 17;10(5):e0004726.
- 21 Andronico A, Dorléans F, Fergé J-L, et al. Real-Time Assessment of Health-Care Requirements During the Zika Virus Epidemic in Martinique. *Am J Epidemiol* 2017; **16**: 1–10.
- 22 SINAN B. SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO. <http://portalsinan.saude.gov.br/>.
- 23 Braga JU, Bressan C, Dalvi APR, et al. Accuracy of Zika virus disease case definition during simultaneous Dengue and Chikungunya epidemics. *PLoS One* 2017; **12**: e0179725.
- 24 Jimenez Corona ME, De la Garza Barroso AL, Rodriguez Martínez JC, et al. Clinical and Epidemiological Characterization of Laboratory-Confirmed Autochthonous Cases of Zika Virus Disease in Mexico. *PLoS Curr* 2016; **8**. doi:10.1371/currents.outbreaks.a2fe1b3d6d71e24ad2b5afe982824053.
- 25 UN. City population by sex, city and city type. <http://data.un.org/>.
- 26 Mallet H, Vial A, Musso D. Bilan de l'épidémie à virus ZIKA en Polynésie Française 2013–2014. *Bul-letin d'Information Sanit Epidemiol Stat* 2015.
- 27 Chan M, Johansson MA. The Incubation Periods of Dengue Viruses. *PLoS One* 2012; **7**: e50972.
- 28 Wilder-Smith A, Gubler DJ, Weaver SC, Monath TP, Heymann DL, Scott TW. Epidemic arboviral diseases: priorities for research and public health. *Lancet Infect Dis* 2017; **17**: e101–6.
- 29 Bogoch II, Brady OJ, Kraemer MU, et al. Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. *Lancet Infect Dis* 2016; **16**: 1237–45.
- 30 Brady OJ, Golding N, Pigott DM, et al. Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasit Vectors* 2014; **7**: 338.
- 31 Kraemer MU, Sinka ME, Duda KA, et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *Elife* 2015; **4**: e08347.
- 32 Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ. Delaying the international spread of pandemic influenza. *PLoS Med* 2006; **3**: e212.
- 33 Roberts MG. Epidemic models with uncertainty in the reproduction number. *J Math Biol* 2013; **66**: 1463–74.
- 34 Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* 2009; **6**: 187–202.
- 35 Yakob L, Kucharski A, Hue S, Edmunds WJ. Low risk of a sexually-transmitted Zika virus outbreak. *Lancet Infect Dis* 2016; **16**: 1100–2.
- 36 Maxian O, Neufeld A, Talis EJ, Childs LM, Blackwood JC. Zika virus dynamics: When does sexual transmission matter? *Epidemics* 2017. doi:10.1016/j.epidem.2017.06.003.
- 37 Jaenisch T, Rosenberger KD, Brito C, Brady O, Brasil P, Marques ET. Risk of microcephaly after Zika virus infection in Brazil, 2015 to 2016. *Bull World Health Organ* 2017; **95**: 191–8.
- 38 Baud D, Gubler DJ, Schaub B, Lanteri MC, Musso D. An update on Zika virus infection. *Lancet* 2017; **390**: 2099–109.
- 39 Stoddard ST, Forshey BM, Morrison AC, et al. House-to-house human movement drives dengue virus transmission. *Proc Natl Acad Sci U S A* 2013; **110**: 994–9.

- 40 Salje H, Lessler J, Maljkovic Berry I, et al. Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size. *Science* 2017; **355**: 1302–6.
- 41 Shutt DP, Manore CA, Pankavich S, Porter AT, Del Valle SY. Estimating the reproductive number, total outbreak size, and reporting rates for Zika epidemics in South and Central America. *Epidemics* 2017; **21**: 63–79.
- 42 Lozier MJ, Burke RM, Lopez J, et al. Differences in prevalence of symptomatic Zika virus infection by age and sex—Puerto Rico, 2016. *J Infect Dis* 2017; published online Dec 6. DOI:10.1093/infdis/jix630.
- 43 WHO. Fifth meeting of the Emergency Committee under the International Health Regulations (2005) regarding microcephaly, other neurological disorders and Zika virus. [http://www.who.int/en/news-room/detail/18-11-2016-fifth-meeting-of-the-emergency-committee-under-the-international-health-regulations-\(2005\)-regarding-microcephaly-other-neurological-disorders-and-zika-virus](http://www.who.int/en/news-room/detail/18-11-2016-fifth-meeting-of-the-emergency-committee-under-the-international-health-regulations-(2005)-regarding-microcephaly-other-neurological-disorders-and-zika-virus)
- 44 Nelson B, Morrison S, Joseph H, et al. Travel Volume to the United States from Countries and U.S. Territories with Local Zika Virus Transmission. *PLoS Curr* 2016; **8**. DOI:10.1371/currents.outbreaks.ac6d0f8c9c35e88825c1a1147697531c.
- 45 Cauchemez S, Ledrans M, Poletto C, et al. Local and regional spread of chikungunya fever in the Americas. *Euro Surveill* 2014; **19**: 20854.
- 46 Baca-Carrasco D, Velasco-Hernández JX. Sex, Mosquitoes and Epidemics: An Evaluation of Zika Disease Dynamics. *Bull Math Biol* 2016; **78**: 2228–42.
- 47 de Oliveira WK, Carmo EH, Henriques CM, et al. Zika Virus Infection and Associated Neurologic Disorders in Brazil. *N Engl J Med* 2017; **376**: 1591–3.
- 48 Carvalho MS, Honorio NA, Garcia LMT, Carvalho LC de S. Aedes aegypti control in urban areas: A systemic approach to a complex dynamic. *PLoS Negl Trop Dis* 2017; **11**: e0005632.
- 49 Codeço CT, Lima AWS, Araújo SC, et al. Surveillance of Aedes aegypti: comparison of house index with four alternative traps. *PLoS Negl Trop Dis* 2015; **9**: e0003475.

Table 1. Summary of the evidence for each population movement model tested on the Zika data. The prior and posterior probabilities were estimated using the ABC-SMC procedure (see SM for further details).

Model of population movements	Gravity (simple) – M_1	Gravity (exponential terms included) – M_2	Radiation– M_3	Flight data– M_4	Combination of flight & radiation– M_5
Prior probability ($\pi(m_y)$)	0.232	0.246	0.224	0.052	0.092
Posterior probability ($P(m_y x)$)	0.001	0.344	0.001	0.001	0.001
Bayes Factor	0.003	1	0.002	0.001	0.001
Evidence for alternative model (and against model m_y)	Very weak evidence of fitting data	Model has best evidence of fitting data	Very weak evidence of fitting data	Very weak evidence of fitting data	Very weak evidence of fitting data

Table 2. Reported and estimated statistics for ZIKV in Latin America and the Caribbean. Reported timing of the peak of ZIKV cases; the model estimate of the peak in ZIKV cases; the estimated number of days each year where $R_0 > 1$; the average value of R_0 throughout the year, the estimated reporting rate of ZIKV cases and the estimated number of ZIKV cases in 2018.

Country	Peak in Data	Peak in model	Days where $R_0(t) > 1^1$	Average $R_0(t)$ during year ¹	% infections that result in a case (reporting rate) ¹	Projected cases in 2018 ¹	Bayesian Posterior Check
Antigua & Barbuda	Sep-16	Dec-16	267 (265, 269)	1.41 (1.36, 1.46)	0.8 (0.6, 1.5)	0 (0, 1)	<0.01
Argentina	Mar-17	Jul-16	122 (121, 123)	1.07 (1.04, 1.11)	0 (0, 0)	6 (2, 15)	>0.99
Aruba	Feb-17	Feb-16	365 (365, 365)	2.41 (2.33, 2.49)	1.3 (0.8, 2.7)	0 (0, 0)	>0.99
Bahamas	Sep-16	Jan-16	254 (254, 255)	2.41 (2.32, 2.48)	0.1 (0.1, 0.4)	0 (0, 0)	>0.99
Barbados	Jan-16	Feb-16	269 (267, 271)	2.13 (2.05, 2.19)	0.4 (0.2, 0.8)	0 (0, 0)	0.22
Belize	Feb-17	Dec-16	238 (236, 239)	1.36 (1.31, 1.4)	0.7 (0.5, 1.3)	3 (0, 13)	>0.99
Bolivia	Feb-17	May-16	256 (254, 259)	1.99 (1.92, 2.06)	0.1 (0.1, 0.3)	0 (0, 0)	>0.99
Brazil	Feb-16	Apr-47	241 (239, 243)	1.99 (1.92, 2.05)	0.7 (0.5, 1)	143 (29, 360)	0.43
Colombia	Dec-16	Jun-16	314 (311, 315)	1.94 (1.87, 2.01)	1.7 (1.3, 2.5)	86 (5, 294)	<0.01
Costa Rica	Sep-16	Jul-16	116 (97, 139)	0.76 (0.74, 0.79)	29.6 (12.5, 55.8)	28 (14, 48)	>0.99
Cuba	Jul-15	Jan-16	260 (259, 261)	2.51 (2.43, 2.6)	0 (0, 0)	0 (0, 0)	<0.01
Curacao	Nov-16	Mar-16	365 (365, 365)	2.22 (2.14, 2.29)	4.7 (2.9, 10)	0 (0, 0)	>0.99
Dominican Republic	May-16	Jun-16	329 (325, 333)	2.21 (2.13, 2.28)	0.3 (0.2, 0.5)	0 (0, 0)	0.06
Ecuador	Jun-16	May-16	130 (130, 131)	1.86 (1.8, 1.92)	0.2 (0.1, 0.5)	0 (0, 1)	>0.99
El Salvador	Dec-16	Nov-16	207 (205, 208)	1.36 (1.31, 1.41)	1.6 (1.2, 2.8)	3 (0, 9)	<0.01
French Guiana	Apr-16	Aug-16	230 (226, 232)	1.16 (1.12, 1.2)	36.9 (22.1, 97.3)	694 (148, 1773)	0.1
Grenada	Jun-16	Jul-16	331 (327, 333)	1.96 (1.9, 2.03)	0.6 (0.4, 1.1)	0 (0, 0)	0.42
Guadeloupe	Jun-16	Jun-16	303 (301, 305)	2.08 (2.01, 2.15)	9.3 (6, 16.9)	0 (0, 0)	0.25
Guatemala	Jan-16	Oct-16	208 (206, 208)	1.59 (1.54, 1.65)	0.5 (0.4, 0.9)	0 (0, 0)	<0.01
Guyana	Jan-16	Aug-16	311 (307, 313)	1.73 (1.67, 1.79)	0.4 (0.3, 0.7)	0 (0, 0)	<0.01
Haiti	Jan-16	Jun-16	295 (293, 296)	2.3 (2.22, 2.38)	0.2 (0.1, 0.3)	0 (0, 0)	<0.01
Honduras	Jan-16	Aug-16	222 (221, 223)	1.85 (1.79, 1.91)	3.7 (2.4, 7.2)	0 (0, 0)	<0.01
Jamaica	Jun-16	Aug-16	269 (268, 271)	1.86 (1.8, 1.92)	1.3 (0.8, 2.5)	0 (0, 0)	<0.01
Martinique	May-16	Aug-16	323 (320, 325)	1.9 (1.83, 1.96)	11.3 (7.3, 20.9)	0 (0, 0)	<0.01
Mexico	Sep-16	Jan-31	141 (139, 142)	1.35 (1.3, 1.39)	0.1 (0.1, 0.1)	5 (2, 9)	0.99
Nicaragua	Jul-16	Aug-16	216 (215, 218)	1.82 (1.75, 1.88)	0.6 (0.4, 1.1)	0 (0, 0)	0.13
Panama	Jan-17	Sep-16	278 (277, 279)	1.69 (1.63, 1.75)	0.5 (0.3, 0.9)	0 (0, 0)	>0.99
Paraguay	Mar-16	Mar-16	295 (293, 297)	2.3 (2.22, 2.37)	0 (0, 0.1)	0 (0, 0)	0.41
Peru	Mar-17	Jun-16	168 (168, 169)	1.6 (1.55, 1.65)	0.2 (0.1, 0.3)	5 (0, 17)	>0.99
Puerto Rico	Aug-16	Jun-16	257 (256, 258)	2.28 (2.2, 2.36)	2.2 (1.5, 3.7)	0 (0, 0)	>0.99
St. Vincent &	Jul-16	Aug-16	322	1.87	0.7 (0.5, 1.3)	0 (0, 0)	0.36

Grenadines			(313, 331)	(1.8, 1.93)			
Suriname	Dec-16	Aug-16	277 (274, 280)	1.6 (1.55, 1.66)	2.4 (1.5, 4.9)	0 (0, 1)	<0.01
Trinidad & Tobago	Aug-16	Sep-16	267 (265, 269)	1.8 (1.73, 1.85)	0.5 (0.3, 0.9)	0 (0, 0)	0.24
US Virgin Islands	Jul-16	Jan-17	251 (247, 255)	1.34 (1.29, 1.38)	21.7 (14, 40.6)	12 (0, 45)	<0.01
Venezuela	Jan-16	Jun-16	271 (268, 276)	2.01 (1.94, 2.08)	0.8 (0.6, 1.1)	1 (0, 2)	<0.01

[†]Estimated median (95% credible intervals).

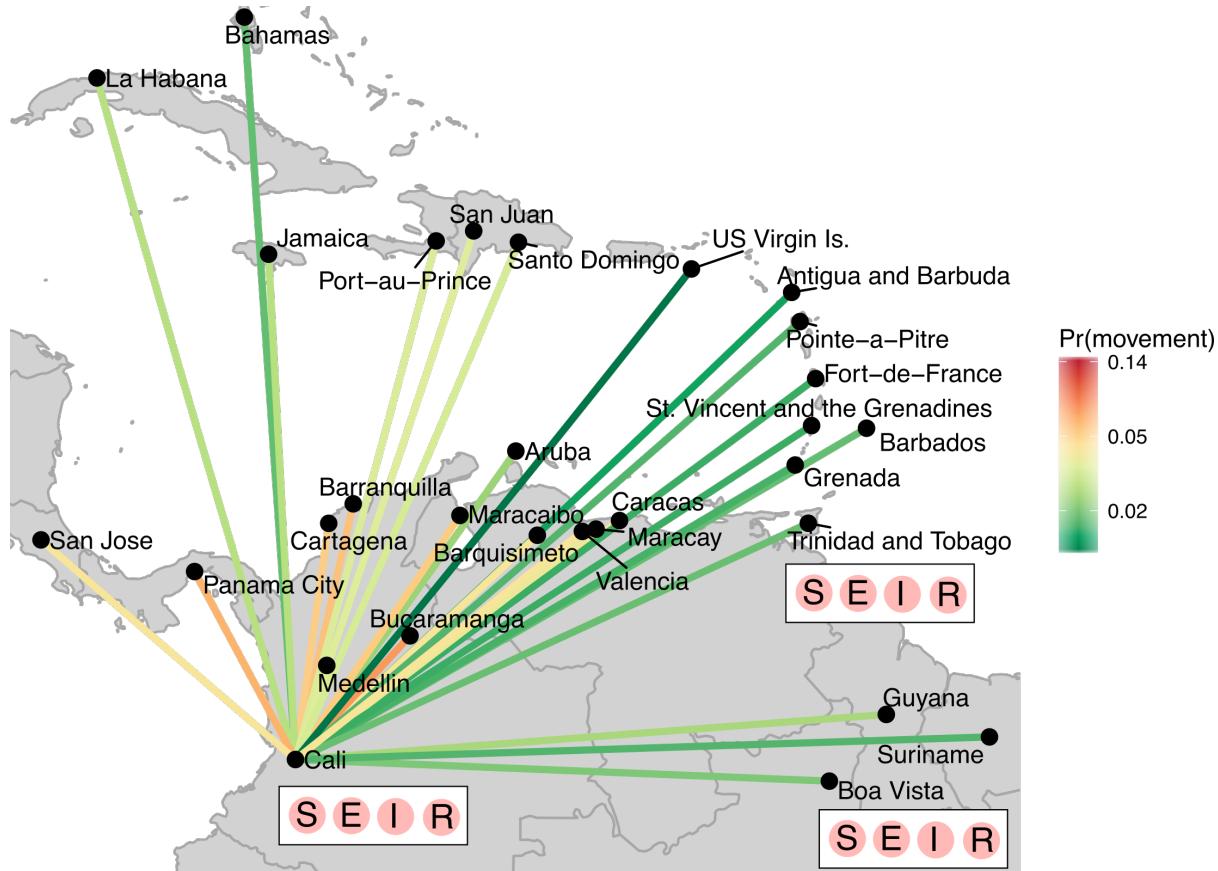


Figure 1. Schematic of the meta-population model structure, that focuses on the northern part of South America and the Caribbean islands. Each city consists of individuals who are assumed to be susceptible (S), pre-infectious (E), infectious (I) or recovered (R) from ZIKV infection. Movement of pre-infectious individuals between cities is modelled assuming different population flows, where a gravity model is illustrated. Movements to cities outside of the plotted area are not illustrated.

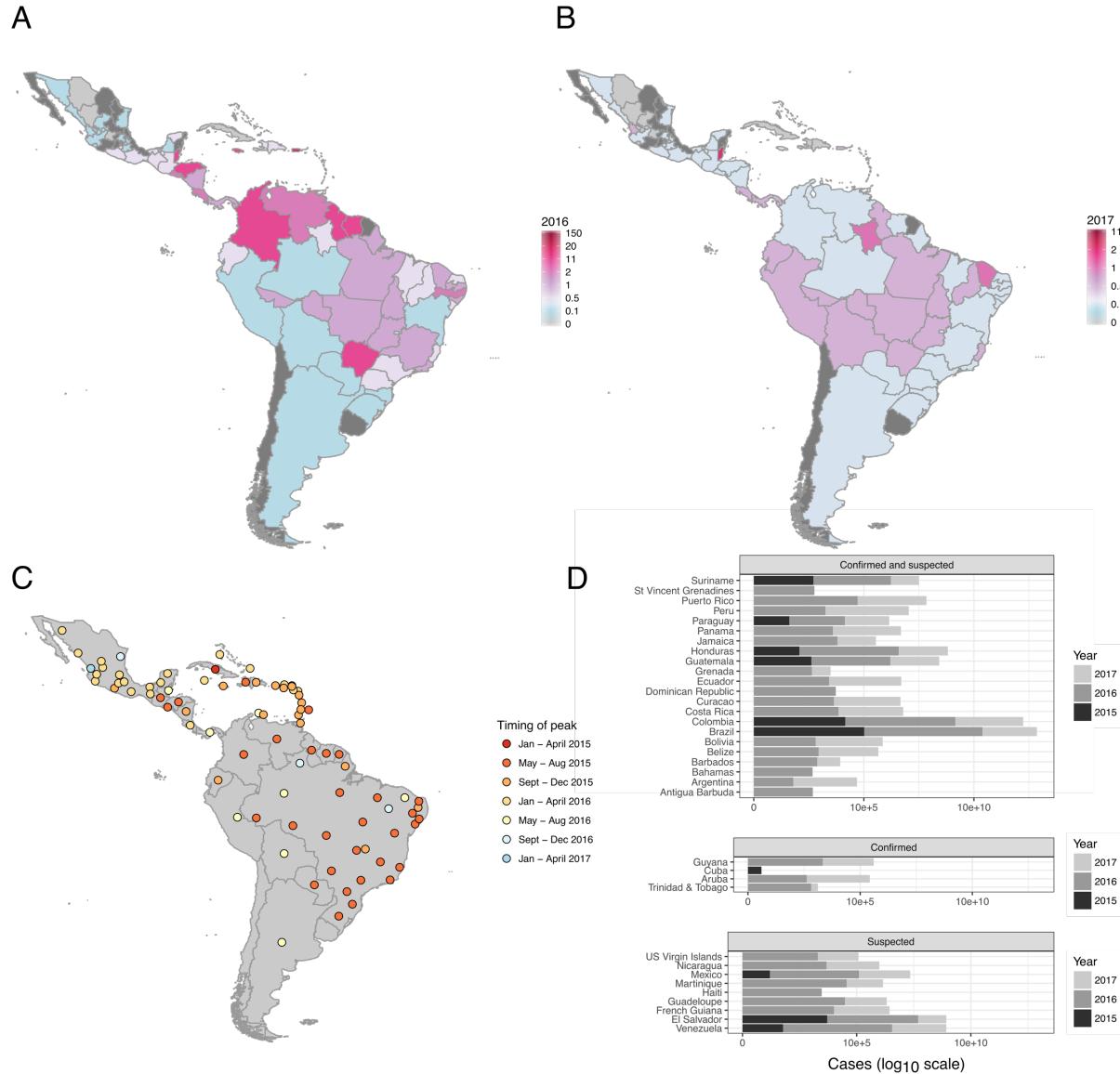


Figure 2. Reported Zika incidence (cases per 1,000) within Latin America for A) 2016 and B) 2017, C) timing of peak incidence and D) total number of cases reported for each country for each calendar year (on a log₁₀ scale), according to the case classifications submitted by each country.

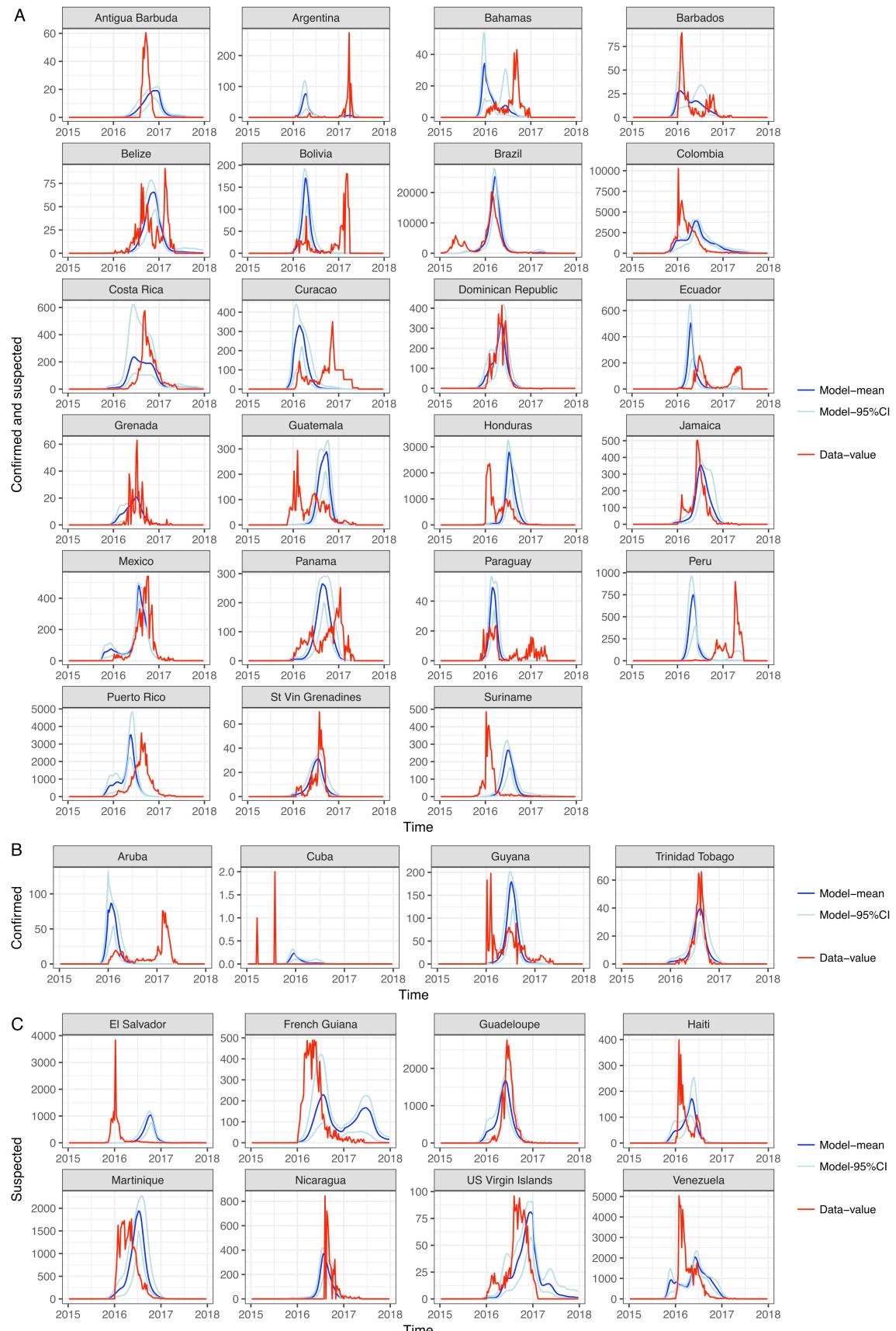


Figure 3. Comparisons of the time-series data and normalised model output for all Latin American countries. The countries are ordered by the type of surveillance data available: Confirmed and suspected, Confirmed, and Suspected cases. To enable comparison we estimate the symptomatic reporting rate for each country and multiply this by the model estimate of the number of infections.

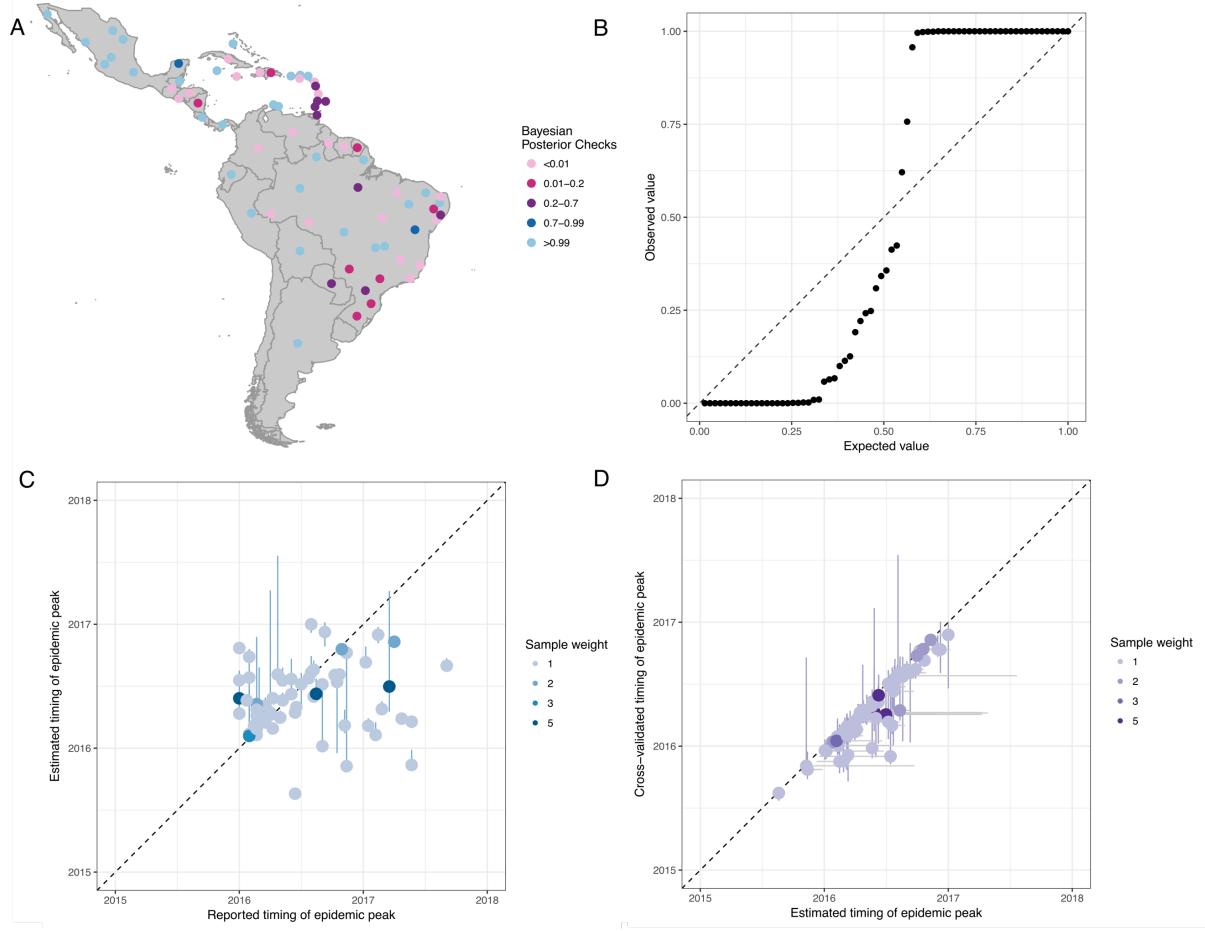


Figure 4. Comparisons of observed and model fit for ZIKV peak incidence in the 31 countries in Latin America. A) Bayesian posterior checks that the estimated peak timing are consistent with the data; values between 0.01-0.99 indicate that the model and data are from the same distribution, **B)** Quantile plot of the Bayesian posterior probabilities **C)** Comparison of the observed timing of the peak and estimated timing of the peak (with 95% CI) **D)** Comparison of the estimated timing of the peak and the cross-validated estimates of peak timing (with 95% CI on the horizontal and vertical).

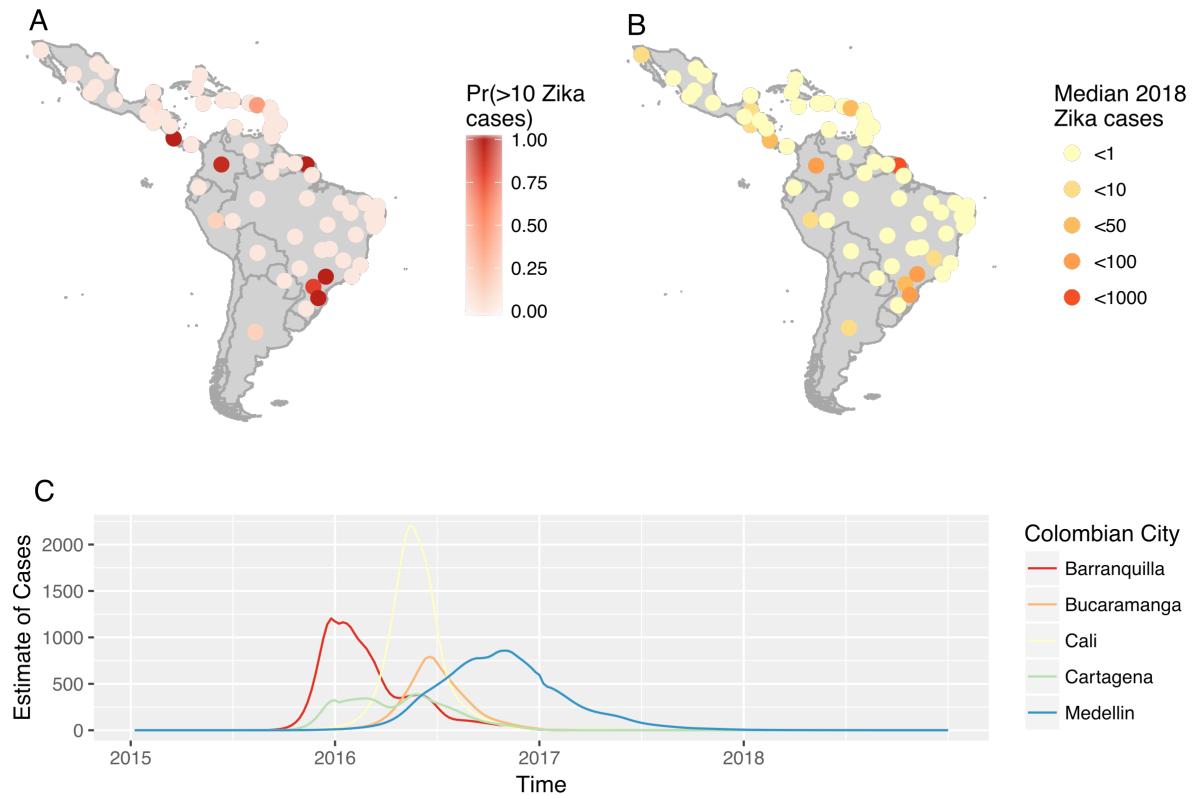


Figure 5. A) The estimated probability of Zika cases in each country (and States in Brazil and Mexico) Probability of more than 10 cases B) Median estimate of Zika cases in 2018 C) The estimated time series of Zika cases within 5 major cities of Colombia.

Supporting Information

Projecting the end of the Zika virus epidemic in Latin America: a modelling analysis

Kathleen M O'Reilly, Rachel Lowe, W John Edmunds, Philippe Mayaud,
Adam Kucharski, Rosalind M Eggo, Sebastian Funk, Deepit Bhatia, Kamran Khan,
Moritz U Kraemer, Annelies Wilder-Smith, Laura C Rodrigues, Patricia Brasil,
Eduardo Massad, Thomas Jaenisch, Simon Cauchemez, Oliver J Brady, Laith Yakob

S1. Compilation of data used in the analysis and selection of cities

Compilation of data

Member states of Latin America and the Caribbean (LAC) have been reporting the incidence of Zika cases to the Pan America Health Organization (PAHO). The data are released as *.pdf* files, along with additional information such as ZIKV in pregnant women and other potential complications. As part of the analysis of the genetics of ZIKV across LAC [1] the Anderson Lab have carefully compiled cases of Zika from the PAHO website [2]. The data are available from January 2015 to August 2017. The process was repeated for PAHO files released up to August 2017. The US Centers for Disease Control also release detailed data for several countries [3]. We first checked the consistency of the incidence curves between PAHO-reported and CDC-reported incidence along with the frequency at which the data were updated on the CDC website. Based upon this, and the need for state-level analysis, only CDC data reported from Mexico were included in the analysis. The weekly number of cases of suspected and confirmed Zika cases for each state in Brazil was obtained from the Brazilian national infectious disease surveillance system (Sistema de Informacao de Agravos de Notificacao - SINAN). Details of this database, case diagnosis methods and changes in definitions have been previously described elsewhere [4].

Selection of cities in the analysis

Cities data were extracted from a publicly available database on world cities, compiled from the UN World Urbanisation Prospects [5, 6]. All cities greater than 750,000 were selected for analysis, as we felt that this balanced the capture of urban locations that may experience Zika and computational time that would be required to estimate parameters from a large mata-population dynamic transmission model. Not all countries within LAC have cities greater than 750,000, and for these countries (typically within the Caribbean) the largest city was selected for analysis. This was repeated for Brazilian and Mexican states. The cities included in the analysis are provided in Table S1, grouped by country. The latitude and longitude of each city was used to infer seasonality curves of Zika transmission, by referencing the global map developed by Brady *et al.* [7] and also used by Bogoch *et al.* [8].

S2. Details of the mathematical model used in the analysis

As described in the main text we assumed that humans were classified by their infection status; susceptible, pre-infections, infections or recovered from ZIKV infection, corresponding to S, E, I, R compartments in a mathematical model of infection. The connectivity between cities is a key component of epidemic spread, and is modelled assuming a meta-population, akin to the *commuter* model described in Keeling and Rohani [9]. Commuters live in one subpopulation (denoted by the *ii* subscript) but travel occasionally to another subpopulation (denoted by the *ij* subscript). From a standard SEIR model we consider the number of individuals of each type in each spatial class and the assumed movement between them:

$$\begin{aligned}
\frac{dS_{ii}}{dt} &= \mu - \beta_i(t)S_{ii}\frac{\sum_j l_{ij}}{\sum_j N_{ij}} - \sum_j l_{ji}S_{ii} + \sum_j r_{ji}S_{ji} - \mu S_{ii} \\
\frac{dS_{ij}}{dt} &= -\beta_i(t)S_{ii}\frac{\sum_j l_{ij}}{\sum_j N_{ij}} + l_{ij}S_{ij} - r_{ij}S_{ij} - \mu S_{ij} \\
\frac{dE_{ii}}{dt} &= \beta_i(t)S_{ii}\frac{\sum_j l_{ij}}{\sum_j N_{ij}} - \alpha E_{ii} - \sum_j l_{ji}E_{ii} + \sum_j r_{ji}E_{ji} - \mu E_{ii} \\
\frac{dE_{ij}}{dt} &= \beta_i(t)S_{ii}\frac{\sum_j l_{ij}}{\sum_j N_{ij}} - \alpha E_{ij} + l_{ij}E_{ij} - r_{ji}E_{ij} - \mu E_{ij} \\
\frac{dI_{ii}}{dt} &= \alpha E_{ii} - \gamma I_{ii} + \sum_j r_{ji}I_{ji} - \mu I_{ii} \\
\frac{dI_{ij}}{dt} &= \alpha E_{ij} - \gamma I_{ij} - r_{ij}I_{ij} - \mu I_{ij} \\
\frac{dR_{ii}}{dt} &= \gamma I_{ii} - \sum_j l_{ji}R_{ii} + \sum_j r_{ji}R_{ji} - \mu R_{ii} \\
\frac{dR_{ij}}{dt} &= \gamma I_{ij} + l_{ij}R_{ij} - r_{ij}R_{ij} - \mu R_{ij} \\
\frac{dN_{ii}}{dt} &= \mu - \sum_j l_{ji}N_{ii} + \sum_j r_{ji}N_{ji} - \mu N_{ii} \\
\frac{dN_{ij}}{dt} &= -l_{ij}N_{ij} + r_{ij}N_{ij} - \mu N_{ij}
\end{aligned}$$

The parameters in the above set of equations are summarized in Table 2, along with the values used in the analysis. To improve computational speed we only model the movement of pre-infectious individuals, as the movement of susceptible individuals does not impact the dynamics but consist of the bulk of movements. Infectious individuals were assumed to not move between cities, due to symptoms, although it is acknowledged that many infectious individuals are asymptomatic. The model was implemented in MatLab (R2017) and R (version 3.3.3). Model code and outputs are available at <https://github.com/kath-o-reilly/Zika-LAC-Outbreaks>.

Movement models fitted to the Zika outbreak

The fit of five movement models were tested against the data on ZIKV outbreaks; i) a gravity-model with no exponential terms; ii) a gravity model with estimated exponential terms; iii) a radiation model; iv) a data-driven approach based on flight data; and v) a model of local radiation and flight

movements. Each model has different assumptions of the connectivity between each city, and will influence the rate of spread of ZIKV between them. There is currently limited information about the likely modes of human movement that influence the spread of ZIKV, but international movements via flights have been cited as a likely mode. For each model we assume that the probability of a pre-infectious individual moving from city i to city j follows the patterns specified in the following equations below. All probabilities are normalised such that $\sum_i \sum_j m_{ij} = 1$.

Model 1: gravity-model with no exponential terms

Movement is driven by city population sizes (n_i and n_j respectively) and the Euclidean distance between them (d_{ij}). The terms μ , ν and γ are exponential terms that affect the strength of size and distance on movement;

$$P(m_{ij}(g_1)) = \frac{n_i^\mu n_j^\nu}{d_{ij}^\gamma}$$

Model 2: gravity model with estimated exponential terms

The exponential terms (g^μ , g^ν and g^γ) of the gravity probability matrix $P(m_{ij}(g_2))$ were estimated from the data alongside the other parameters.

Model 3: radiation model

The radiation model was developed to account for local population movements being influenced by the need to travel until required resources (eg. employment) are met. The parameter s_{ij} described the population size within the circle defined by the locations of cities i and j (with coordinates $\{x_i, y_i\}$ and $\{x_j, y_j\}$ respectively);

$$P(m_{ij}(r)) = \frac{n_i n_j}{(n_i + s_{ij})(n_i + n_j + s_{ij})}$$

Population sizes were extracted from WorldPop, and to save on computational time if the $d_{ij} > 6000$ km $P(m_{ij}(r))$ was assigned a value of 10^{-5} .

Model 4: flight data Data from selected airports that correspond to the selected cities were extracted from the Global database of the International Air Transport Association (www.iata.org). Some cities (eg. Sao Paulo) have multiple airports and data from all airports associated with the city were extracted. For smaller (eg. Caribbean) countries all airports within the country were used in the analysis. The data of the total number of registered tickets between cities from January 2015 to December 2016 were aggregated to provide an estimate of the probability of moving from city i to city j , $P(m_{ij}(f))$. Movements in 2017 were assumed to follow the same patterns.

Model 5: radiation and flight data

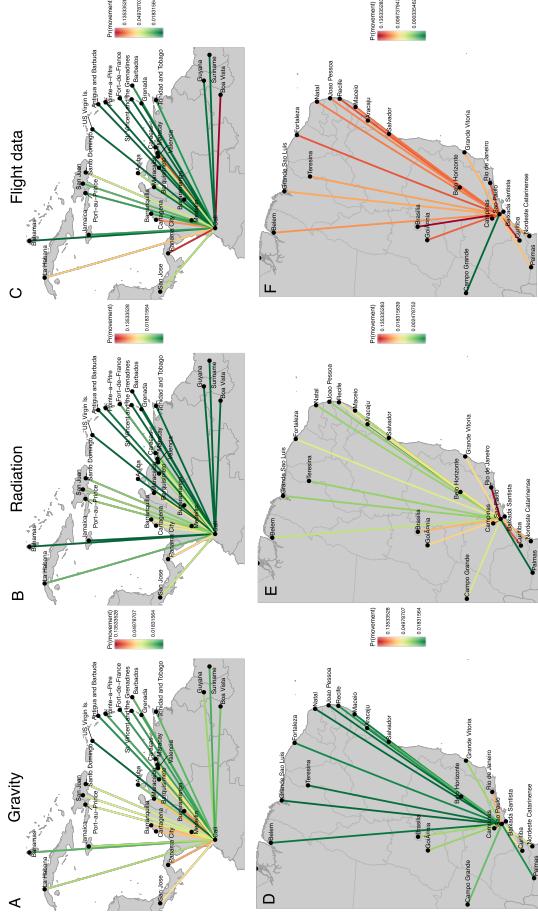
The radiation ($P(m_{ij}(r))$) and flight ($P(m_{ij}(f))$) matrices were combined together assuming a mixture model with parameter η that describes the contribution of the radiation matrix to movement;

$$P(m_{ij}(a)) = \eta P(m_{ij}(r)) + (1 - \eta)P(m_{ij}(f))$$

Comparison of the different movement models

Figure S2 shows the effect of the three main (gravity, radiation and flight) movement models on movement out of Cali (Colombia) and Sao Paulo (Brazil). Both the gravity and radiation models predict that movement mostly occurs between cities within a short distance, and especially among large local cities for the radiation model. The flight data illustrates a different pattern, consisting of a higher probability of movements to large distant cities associated with tourism or as important

commerce hubs.



S2 Figure. Illustration of the three different movement models tested in the analysis; (A, D) gravity model, (B, E) radiation model and (C, F) flight data. The top row shows the estimated probability of movement out from Cali (Colombia) and the lower row shows the estimated probability of movement out from Sao Paulo (Brazil).

Additional details of using ABC-SMC to fit the data to the models

Let $m\{1, \dots, M\}$ be a model parameter that specifies the movement model, where M is the total number of models. The model-specific parameters are denoted by $\theta(m) = (\theta(m)^{(1)}, \dots, \theta(m)^{(k_m)})$, $m = 1, \dots, M$, where k_m denotes the number of parameters in model m . Following the algorithm given in Toni et al [16], ABC was used to approximate the posterior distribution $\pi(\theta|x)$ based on the distance between the model outputs and the data being below a specified tolerance value $d(x_0, x^*) \leq \epsilon$. The distance function is a comparison of the timing of the peak of ZIKV epidemics within each of the geographical units in the analysis (countries of Latin America and the Caribbean, and States within Brazil and Mexico) and the modelled cities, ie. $X_{p,i}$ describes the reported peak for each geographical unit in the data and $Z_{p,i}$ describes the estimated peak from the mathematical modelling (where city timeseries Z_j were summed to provide Z_i). To account for multiple cities being present within some geographical units, the model and data were weighted by the number of cities (n_i) within each geographical unit $d(x_0, x^*) = \sum (X_{p,i} - Z_{p,i})^2 / n_i^2$.

We start by running a set of particles ($n = 100$) where each particle simulates the ZIKV epidemic assuming model m . The parameters of each model are initially selected from the priors specified in Table 2, m^* from $\pi(m)$ and θ^* from $\pi(\theta(m^*))$. 1000 iterations of each particle were run, where parameter sets were randomly picked from the range in Table 2 and the accepted parameter sets were used to inform the multivariate normal distribution used in the second round. For the second round, 100 particles were again selected, but particles were sampled using multinomial distribution

with probabilities from the previous round. Parameter sets of a selected model (ie. θ^{**}) were sampled from a perturbation kernel based on the accepted parameter sets from the previous population and the estimated covariance matrix ($\theta^{**} = K_t(\theta|\theta^*)$). An ABC-MCMC algorithm was used to approximate the posterior distribution, ie. θ^{**} was accepted with probability $\alpha = \min(1, \frac{\pi(\theta^{**})q(\theta_t|\theta^{**})}{\pi(\theta_t)q(\theta^{**}|\theta_t)})$ when $d(x_0, x^*) \leq \epsilon$, otherwise $\theta_{t+1} = \theta_t$. 5,000 iterations of each model were run. The mean and 95% credible intervals (CrI) of each parameter were estimated from the particles where the preferred model was selected.

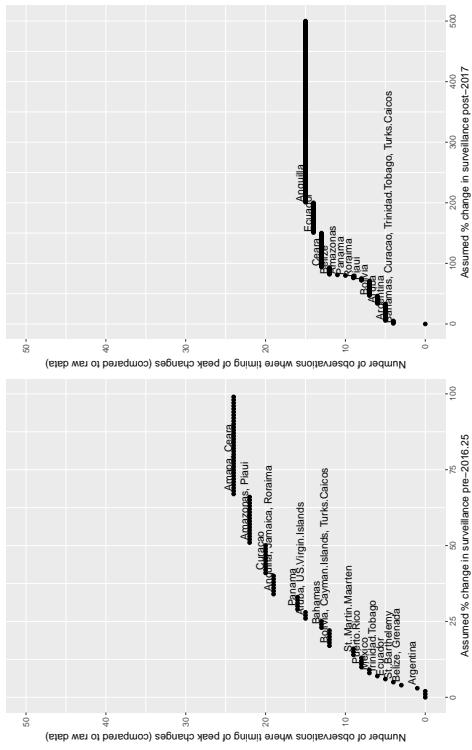
Bayes Factors were used for model selection. From the ABC-SMC we have an approximation of the marginal posterior distribution, based on the particles with accepted parameter sets. Using the particles from the second population Bayes Factors were estimated using $BF_{12} = \frac{P(m_1|x)}{P(m_1)F(m_2)}$. The best fitting model was based on interpretation of the Bayes Factors provided by Kass and Raftery [17].

S3. Testing for the impact of surveillance sensitivity on the estimated timing of the peak in incidence

A concern with the data quality of the Zika epidemic is that surveillance for ZIKV cases may have substantially increased from January 2016 (due to an increasing awareness about the epidemic) and may have reduced after December 2016 (due to reduced reporting of cases and less media coverage on the epidemic). Moreover, Zika-specific surveillance systems in LAC became operational in December 2015, and PAHO reported incidence only begins in January 2016 despite cases being reported in several countries prior to this time. To test the impact of changes in surveillance within each country we applied a scaling factor to each time series for observations prior to March 2017 and after December 2016. For each country time series we have the weekly number of reported Zika cases ($\mathbf{X} = \{X_t, t = 1, \dots, T\}$) and we apply scaling factors to produce an updated time series ($\mathbf{Y} = \{Y_t, t = 1, \dots, T\}$):

$$Y_t = \begin{cases} aX_t & \text{if } t < 2016.25 \\ X_t & \text{if } 2016.25 \leq t < 2017 \\ bX_t & \text{if } t \geq 2017 \end{cases}$$

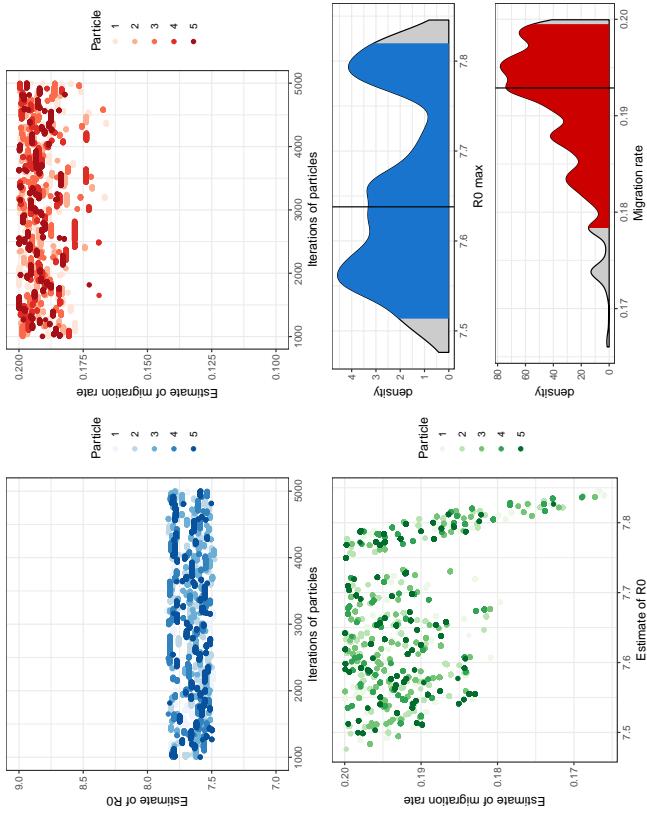
We then identified the number of countries where the peak incidence would change as a result of the change in reporting. We varied the scaling factor within plausible value to determine the effects on estimated timing of the peak. Figure S1 shows that surveillance prior to March 2016 would need to be 25% or less than 2016 intensity for 10 or more observations to have a change in the estimated peak, making the peak incidence quite robust to changes in surveillance. The peak in incidence would change for 13 countries if surveillance had reduced in 2017, and increasing surveillance would affect the peak estimate for 2 countries, again illustrating that peak incidence is a useful summary statistic in this setting.



S3 Figure. Impact of assuming a step-change in surveillance (left) prior to March 2016 and (right) after December 2016, on the estimated peak in ZIKV cases.

S4. Results from the parameter estimation

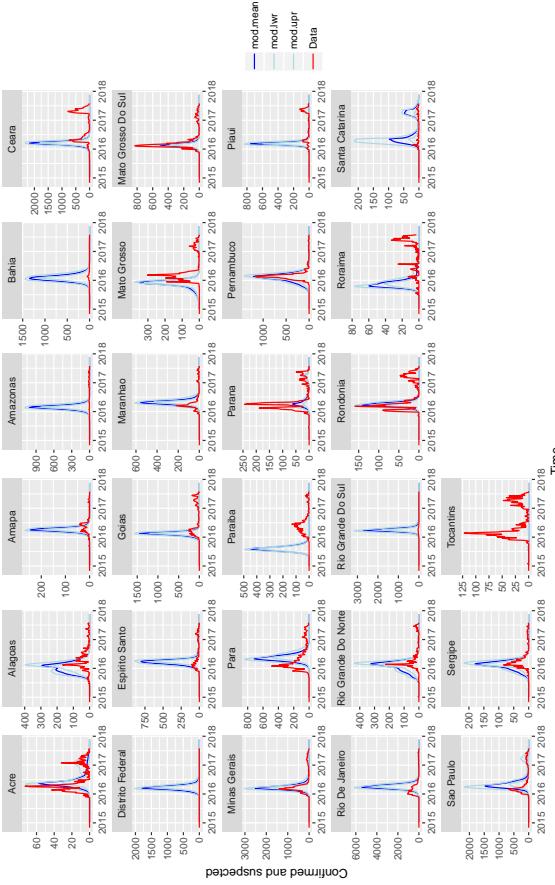
Modelling of the first population of MCMC particles assume that each model had an equal probability of fitting the data. The posterior probability that model m would be selected was used as the prior for the second population, and the probability is shown in Table 1 of the main text. A selection ($n=5$) of particles and their MCMC chains from the second population is shown in Figure S3. To reduce auto-correlation the chains were thinned to every tenth iteration. We illustrate the maximal value of R_0 and the estimated leave rate from the migration matrix. Whilst the chains seem quite stationary, a correlation plot between R_0 and the leave rate show a strong negative correlation when R_0 reaches high values. Summary values of the parameters are provided in Table 3.



S3 Figure. MCMC output of the parameter estimation from 5 particles where the gravity model was selected.

S5. Analysis of data and model comparisons for Brazilian States

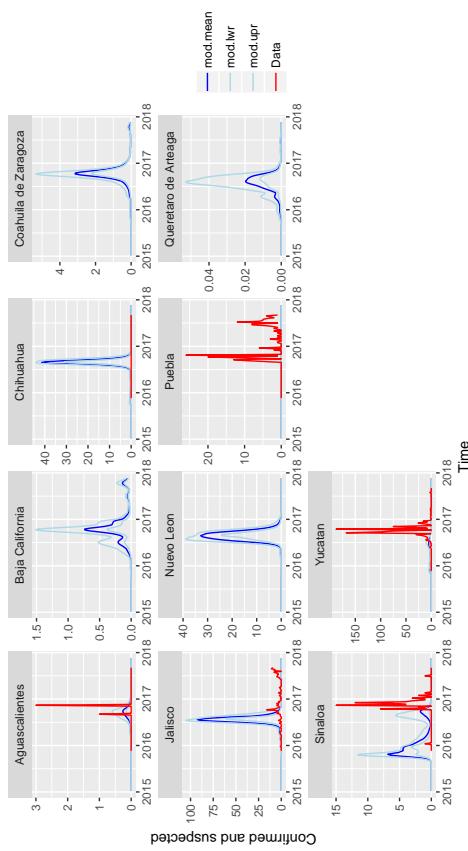
Data were available on suspected and confirmed Zika cases within Brazil at a State level, enabling a state-level analysis of the epidemic. All cities greater than 750,000 were included in the analysis. For states that do not have cities of this size, the largest city was also included in the model (Figure S4).



S5 Figure. Comparison of reported Zika cases to the median (and 95% CI) estimate of cases from the mathematical model.

S6. Analysis of data and model comparisons for Mexican States

Data were available on suspected and confirmed Zika cases within Mexico at a State level, enabling a state-level analysis of the epidemic. All cities greater than 750,000 were included in the analysis, and because cases reported outside of these states were low the analysis focussed on only these states (Figure S5).



S6 Figure. Comparison of reported Zika cases to the median (and 95% CI) estimate of cases from the mathematical model.

References

- [1] Grubaugh, N. D. *et al.* Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* **546**, 401–405 (2017). URL <http://www.nature.com/doifinder/10.1038/nature22400>.
- [2] Andersen, K. PAHO Zika numbers - Andersen Lab (2017). URL <https://andersen-lab.com/paho-zika-cases/>.
- [3] Centers for Disease Control and Prevention (CDC). GitHub repository - cdcepi/zika (2017).
- [4] de Oliveira, W. K. *et al.* Infection-related microcephaly after the 2015 and 2016 Zika virus outbreaks in Brazil: a surveillance-based analysis. *Lancet (London, England)* **390**, 861–870 (2017). URL <http://linkinghub.elsevier.com/retrieve/pii/S0140673617313685http://www.ncbi.nlm.nih.gov/pubmed/28647172>.
- [5] UN. World Urbanization Prospects - Population Division - United Nations. URL <https://esa.un.org/unpd/wup/>.
- [6] Nordpil. World database of large urban areas, 1950-2050. URL <https://nordpil.com/resources/world-database-of-large-cities/>.
- [7] Brady, O. J. *et al.* Global temperature constraints on Aedes aegypti and Ae. albopictus persistence and competence for dengue virus transmission. *Parasites & Vectors* **7**, 338 (2014). URL <http://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-7-338>.
- [8] Bogaoh, I. I. *et al.* Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. *The Lancet Infectious Diseases* **16**, 1237–1245 (2016). URL [http://dx.doi.org/10.1016/S1473-3099\(16\)30270-5](http://dx.doi.org/10.1016/S1473-3099(16)30270-5).
- [9] Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, 2008).
- [10] Kucharski, A. J. *et al.* Transmission Dynamics of Zika Virus in Island Populations: A Modelling Analysis of the 2013 and 2014 French Polynesia Outbreak. *PLoS Neglected Tropical Diseases* **10**, 1–15 (2016). URL <http://dx.doi.org/10.1371/journal.pntd.0004726>.
- [11] Campos, G. S., Bandeira, A. C. & Sardi, S. I. Zika Virus Outbreak, Bahia, Brazil. *Emerging Infectious Diseases* **21**, 1885–1886 (2015). URL [http://www.ncbi.nlm.nih.gov/pubmed/26401719http://www.ncbi.nlm.nih.gov/article/PMC4693454http://www.ncbi.nlm.nih.gov/article/21/10/15-0847\[.\]article.htm](http://www.ncbi.nlm.nih.gov/pubmed/26401719http://www.ncbi.nlm.nih.gov/article/PMC4693454http://www.ncbi.nlm.nih.gov/article/21/10/15-0847[.]article.htm).
- [12] Lessler, J. *et al.* Times to key events in the course of zika infection and their implications for surveillance: A systematic review and pooled analysis. *bioRxiv* (2016). URL <http://www.biorxiv.org/content/early/2016/03/02/041913.html>.
- [13] Venturi, G. *et al.* An autochthonous case of Zika due to possible sexual transmission, Florence, Italy, 2014. *Eurosurveillance* **21**, 30148 (2016). URL <http://www.ncbi.nlm.nih.gov/pubmed/26939607http://www.eurosurveillance.org/viewarticle.aspx?ArticleId=21305>.

- [14] Bank, W. Life expectancy at birth (2017). URL <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>.
- [15] Gog, J. R. *et al.* Spatial Transmission of 2009 Pandemic Influenza in the US. *PLoS Computational Biology* **10**, e1003635 (2014). URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4052844/> <http://dx.plos.org/10.1371/journal.pcbi.1003635>.
- [16] Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. H. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society, Interface* **6**, 187–202 (2009). URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2658655>.
- [17] Kass, R. E. & Raftery, A. E. Bayes Factor (1995).

Table 1: Country and State (in Brazil and Mexico) of cities included in the analysis

Country	State	Cities included in analysis
Antigua and Barbuda		Buenos Aires, Cordoba, Mendoza, Rosario, San Miguel de Tucuman
Argentina		
Aruba		
Bahamas		
Barbados		
Belize		
Bolivia	Santa Cruz	
Brazil	Acre	Rio Branco
Brazil	Alagoas	Maceio
Brazil	Amapa	Macapa
Brazil	Amazonas	Manaus
Brazil	Bahia	Salvador
Brazil	Ceara	Fortaleza
Brazil	Districto Federal	Brasilia
Brazil	Espirito Santo	Grande Victoria
Brazil	Goias	Goiania
Brazil	Maranhao	Grande Sao Luis
Brazil	Mato Grosso	Cuiaba
Brazil	Mato Grosso do Sul	Campo Grande
Brazil	Minas Gerais	Belo Horizonte
Brazil	Para	Belen
Brazil	Paralba	Joao Pessoa
Brazil	Parana	Curitiba
Brazil	Pernambuco	Recife
Brazil	Piaui	Teresina
Brazil	Rio de Janeiro	Rio de Janeiro
Brazil	Rio Grande do Norte	Natal
Brazil	Rio Grande do Sul	Ponto Alegre
Brazil	Rondonia	Ponto Velho
Brazil	Roraima	Boa Vista
Brazil	Santa Catarina	Florianopolis, Nordeste Catarinense
Brazil	Sao Paulo	Baixada Santista, Campinas, Sao Paulo
Brazil	Sergipe	Aracaju
Brazil	Tocantins	Palmas
Colombia	Baranquilla, Bucaramanga, Cali, Cartagena, Medellin	
Costa Rica		
Cuba		
Curacao		
Dominican Republic		
Ecuador		
El Salvador		
French Guiana		
Grenada		
Guadeloupe		
Guatemala		
Guyana		
Haiti		
Honduras		
Jamaica		
Martinique		
Mexico		
Nicaragua		
Panama		
Paraguay		
Peru		
Puerto Rico		
Saint Vincent and the Grenadines		
Suriname		
Trinidad and Tobago		
United States Virgin Islands		
Venezuela		
		Charlotteville, Caracas, Maracaibo, Maracay, Valencia
		Baquisimeto

Parameter	Definition	Values used in analysis	Source
	<i>Main parameters</i>		
$\beta_i(t)$	Time-varying transmission rate	Varied across settings	Estimated from data
$1/\alpha$	Intrinsic incubation period	5 days	[10] [11] [12] [13]
$1/\gamma$	Human infectious period	20 days	[10] [11] [12] [13]
$1/\mu$	Human life expectancy	75 years	[14]
l_{ij}	Rate that humans leave city j and commute to city i	$P(m_{ij})\kappa_m$	Estimated from data
r_{ij}	Rate of return from i to j	0.001 per day	[9]
	<i>Model-specific parameters</i>	<i>Range of values tested in ABC-SMC algorithm</i>	
κ_g , κ_r and κ_f	Scaling parameters for gravity, radiation and flight model	All varied between 10^{-10} and 10^{-1}	
g^μ , g^ν and g^γ	Exponents of the gravity model	Varied from $[0.01, 1]$ and $[0.1, 1]$ respectively	[15]

Table 2: Parameters used in the model.

Parameters estimated in the model	Median value (95% CrI)
$R_{0,max}$	7.00 (5.61, 8.51)
leave rate	0.03 (0.0003, 0.19)
g^μ	2.1 (2.0, 2.2)
g^ν	0.36 (0.1, 0.48)
g^γ	0.12 (0.5, 0.90)

Table 3: Parameter estimates (and 95% credible intervals) from the gravity model fitted to the ZIKV data and mathematical model