Model-based projections of Zika virus infections in childbearing women in the Americas

T. Alex Perkins[1*], Amir S. Siraj[1], Corrine Warren Ruktanonchai[2], Moritz U.G. Kraemer[3], Andrew J. Tatem[2,4]

[1] Department of Biological Sciences and Eck Institute for Global Health, 100 Galvin Hall, University of Notre Dame, Notre Dame, IN 46556 USA
[2] WorldPop, Department of Geography and Environment, University of Southampton, Southampton, SO17 1BJ, UK
[3] Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK
[4] Flowminder Foundation, Stockholm, Sweden

[*] Author for correspondence: taperkins@nd.edu

Zika virus is a mosquito-borne pathogen that is rapidly spreading across the Americas. Due to associations between Zika virus infection and a range of fetal maladies[1,2], the epidemic trajectory of this viral infection poses a significant concern for the nearly 15 million children born in the Americas each year. Ascertaining the portion of this population that is truly at risk is an important priority. One recent estimate[3] suggested that 5.42 million childbearing women live in areas of the Americas that are suitable for Zika occurrence. To improve on that estimate, which did not take into account the protective effects of herd immunity, we developed a new approach that combines classic results from epidemiological theory with seroprevalence data and highly spatially resolved data about drivers of transmission to make location-specific projections of epidemic attack rates. Our results suggest that 1.65 (1.45–2.06) million childbearing women and 93.4 (81.6–117.1) million people in total could become infected before the first wave of the epidemic concludes. Based on current estimates of rates of adverse fetal outcomes among infected women[2,4,5], these results suggest that tens of thousands of pregnancies could be negatively impacted by the first wave of the epidemic. These projections constitute a revised upper limit of populations at risk in the current Zika epidemic, and our approach offers a new way to make rapid assessments of the threat posed by emerging infectious diseases more generally.

On February 1, 2016, the World Health Organization (WHO) designated the ongoing Zika virus epidemic in the Americas as a Public Health Emergency of International Concern (PHEIC), defined as an "extraordinary event" that "potentially require[s] a coordinated international response"[6]. This declaration acknowledges the high potential for Zika to establish across the Americas given that its dominant vector, the *Aedes aegypti* mosquito, is endophilic and occupies an exceptionally broad range[7]. Concern underlying this rare WHO declaration also stems from an association between Zika virus infection in pregnant women and a range of adverse fetal outcomes[2], most notably congenital microcephaly[1]. As of April 28, 2016, there were 1,286 confirmed cases of microcephaly associated with Zika virus infection in five countries[8], and there is widespread concern that these numbers could increase further as the virus continues to spread across the Americas[9].

A number of uncertainties surround the future of the Zika epidemic in the Americas, particularly questions about how many women may be at risk of having children with congenital microcephaly and other adverse outcomes associated with Zika virus infection[10]. Of women who become infected with Zika virus during a vulnerable stage of their pregnancy, evidence is emerging that 1-13% may go on to develop congenital microcephaly[2,4,5]. However, the number of women who become infected with Zika virus during that timeframe is difficult to ascertain. One recent study[3] estimated that 5.42 million births occurred in 2015 in regions of the Americas with "suitability" for Zika "occurrence." Such estimates come with many caveats though, as they rely on a relatively limited number of reported cases and apply a method based on equilibrium assumptions to a situation involving active range expansion[11]. Most importantly, the estimate of 5.42 million births[3] reflects the total population within a demarcated area and does not take into account that large fractions of the populations in those areas may remain uninfected due to herd immunity and other factors[12,13].

To quantify the potential magnitude of the ongoing Zika epidemic in terms of people who realistically might become infected, we formulated and applied a method for projecting location-specific epidemic attack rates on highly spatially resolved human demographic projections[14]. The central concept behind our approach is that of the "first-wave" epidemic. Zika and other mosquito-borne viruses have been known to exhibit explosive outbreaks, infecting as much as 75% of a population in a single year[15]. Classical epidemiological theory predicts that some proportion of a population will remain uninfected during an epidemic, because herd immunity eventually causes the epidemic to burn out[12]. A related prediction of this theory is that the proportion infected prior to epidemic burnout (i.e., the epidemic attack rate) has a one-to-one relationship with the basic reproduction number, R_0[13]. The latter quantity has a well-known mechanistic formulation for mosquito-borne pathogens[16] that accommodates the effects of environmental drivers on transmission[17,18]. For example, the incubation periods of dengue viruses in the *Ae. aegypti* mosquitoes that transmit Zika virus have an empirically derived relationship with temperature[18], which can in turn be used to inform calculations of R_0. Together with similar relationships for other transmission parameters, it is possible to characterize R_0, a fundamental measure of transmission potential, as a function of local environmental conditions.

We leveraged these classic results from epidemiological theory to first perform highly spatially resolved calculations of R_0 and then to translate those calculations into location-specific projections of first-wave epidemic attack rates (Fig. 1a). Because Zika-specific values of transmission parameters are largely unknown at present but may be well approximated by dengue-specific values[19], we used some parameter values for dengue virus in our R_0 calculations. We furthermore calibrated our attack rate projections to match empirically estimated attack rates from 12 chikungunya epidemics and one Zika epidemic in naïve populations (Extended Data Table 1). This step afforded us the flexibility to enhance the realism of the model with respect to firmly established but poorly quantified associations between human-mosquito contact and economic prosperity[20]. In doing so, one departure from the classic relationship between R_0 and attack rate that we made was to rescale R_0 by an exponent $\alpha \in (0,1]$ to allow for better correspondence with observed attack rates. Although there is no theoretical justification for this or any other particular scaling relationship, it is consistent with theoretical expectations[21] that attack rates should be lower in populations with equal R_0 values but more heterogeneous contact

patterns, which are typical for transmission by *Ae. aegypti*[22]. To provide a point of reference for our model-based approach, we also fitted a statistical description of the 13 seroprevalence estimates as a function of the environmental drivers that we considered. For both approaches, we applied their respective location-specific attack rate projections to demographic projections on a 5x5 km grid across Latin America and the Caribbean to obtain expected numbers of infections in the overall population and among childbearing women in particular (Fig. 1c). All such calculations were performed for 1,000 Monte Carlo samples of model parameters.

In total, our median projection suggests that as many as 93.4 (range: 81.6-117.1) million people in Latin America and the Caribbean could become infected during the first wave of the epidemic (Fig. 1b). To place this number into context, we refer to an estimate[23] that 53.8 (40.0–71.8) million dengue infections occurred in this region in 2010 alone. Our projections of nearly double this number for Zika are not surprising, given that there is extensive immunity to dengue but not Zika in this region and given that it would likely take longer than a year for the first wave of the epidemic to conclude in all locations within this region. At the country level, we project that Brazil will have the largest total number of infections by more than double that of any other country, due to a combination of its size and suitability for transmission. Island countries in the Caribbean are projected to experience the highest nationally averaged attack rates, with 7 of the highest 10 values projected for countries including Aruba, Haiti, and Cuba. This projection is consistent with a frequent history of arbovirus outbreaks on islands[24] and may owe to the uniformity of environmental conditions on the portions of islands where people tend to live. In more heterogeneous regions, the 5x5 km spatial resolution of our maps allows for nuanced projections for areas of interest to local stakeholders (Fig. 1d,e).

Among childbearing women, our median projection suggests that there could be as many as 1.65 (range: 1.45–2.06) million infections in Latin America and the Caribbean before the first wave of the epidemic concludes (Fig. 1b). Assuming that birth rates are temporally constant, our projections are robust to uncertainty about the timing of local epidemics and the timeframe of the first wave of the epidemic, because they are based on cumulative proportions infected. These projections can also be used to postulate numbers at risk of microcephaly by multiplying them by the fraction of a year in which a pregnant woman is susceptible to developing microcephaly (e.g., multiply by 1/4 in the case of first-trimester susceptibility). We also note that there were some discrepancies in our projections in terms of the rank order of countries experiencing the most infections among childbearing women versus the population as a whole. In particular, Cuba was 5th in terms of projected infections in the overall population but 12th in terms of infections among childbearing women due to its low birth rate compared to other countries in the Americas[25]. Such discrepancies are also likely to exist subnationally[26], and their elucidation should be a priority for future work.

By accounting for uncertainty distributions for each of the key drivers of our model (Fig. 2a-e), we found that uncertainty distributions for infections across the region as a whole and by country were often multimodal (Fig. 2f-o) due to uncertainty in the shape of the relationship between mosquito-human contact and the local economic index that we considered (Fig. 2d). Summing our projections across Latin America and the Caribbean revealed variation that was modest, in the sense that none of our 1,000 Monte Carlo samples resulted in fewer than 81 million infections overall and 1.4 million among childbearing women (Fig. 2f,k). There are many

reasons that even these numbers could be overestimates though. Our projections are conditioned on a local epidemic taking place in each 5x5 km grid cell in the region, which is unlikely to happen given dispersal limitation, stochastic fadeout, geographic mismatches in seasonality, and other factors. Therefore, it is most appropriate to interpret our projections as either a plausible worst-case scenario or an expectation of local epidemic size conditional on there being a local epidemic in the first place.

Although our approach was very much rooted in mechanistic models from epidemiological theory, two critical steps in our method involved fitting curves to describe theoretically motivated but heretofore unknown relationships: an association between mosquito-human contact and economic prosperity (Fig. 2d), and a scaling relationship between R_0 and attack rates (Fig. 2e). Allowing these relationships to be informed by local seroprevalence estimates (Extended Data Table 1) left open the question of the extent to which our projections were informed by the mechanistic assumptions of the model versus statistical fits to the seroprevalence estimates that we used. On the one hand, an alternative statistical approach accounted for much more variation in seroprevalence estimates (R^2=0.89) than did the model-based approach (R^2=0.32). On the other hand, the statistical approach offered a dichotomous set of projections about numbers of infections outside the context of the data to which it was fitted: either everyone will become infected or very few people will (Fig. 3). Relationships between attack rates and predictor variables inferred by the statistical approach (Fig. 4d-i) were also implausible: a narrow temperature range in which attack rates increase sharply towards 100% (Fig. 4d-h), and a reversal of economic effects whereby wealthy populations experience higher attack rates than poor populations when mosquito occurrence probabilities are high (Fig. 4f,i). By contrast, the model-based approach yielded more moderate attack rate projections overall (Fig. 2f vs 3a) in which temperature, economic prosperity, and mosquito occurrence probability all had plausible relationships with attack rates (Fig. 4a-c).

In conclusion, our model-based approach offers a unique way to leverage a variety of spatially detailed data products[7,14,27,28] to make a priori projections of attack rates and infections that could be experienced in the first wave of the ongoing Zika epidemic. Projections such as these have an important role to play in the early stages of an epidemic, when planning for surveillance and outbreak response is actively underway both internationally and locally[9]. At the same time, it is important for consumers of this information to be aware of uncertainties in these and other projections, which often exceed the amount of uncertainty that can be identified a priori[29]. Likewise, following up on these projections in the aftermath of the epidemic—by comparing against projections made with alternative models and additional serological surveys[30]—will provide an exceptional opportunity to enhance capabilities to anticipate the severity of future epidemic threats.

METHODS

Data sources and processing

Human demography

To estimate the annual numbers of pregnancies per 1x1 km grid cell in 2015, methods developed by the WorldPop project (www.worldpop.org[25,31]) were adapted for the Americas region. High-resolution estimates of population counts per 100x100 m grid cell for 2015 were recently constructed for Latin American, Asian, and African countries[14,32]. With consistent subnational

data on sex and age structures, as well as subnational age-specific fertility rate data across the Americas currently unavailable for fully replicating the approaches of Tatem et al.[31], national level adjustments were made to construct pregnancy and birth counts. Data on estimated total numbers of births[33] and pregnancies[31] occurring annually in 2012 were assembled for all Latin American study countries, as well as births in 2015[33]. As no 2015 pregnancy estimates existed at the time of writing, the ratios of births to pregnancies for each country in the Americas were calculated using 2011 and 2012 estimates, and these were then applied to the 2015 births numbers to obtain 2015 estimates of annual pregnancy numbers per-country. This made the assumption that per-country births-to-pregnancies ratios remained the same in 2015 as they were in 2011 and 2012. The 100x100 m gridded population totals were aggregated to 1x1 km spatial resolution, and the per-country totals were linearly adjusted to match the 2015 pregnancy estimates.

Temperature

We used interpolated meteorological station temperature data from the 1950-2000 period at 5x5 km spatial resolution, processed to create climatological monthly averages that represent "typical" conditions (www.worldclim.org[27]).

Aedes aegypti occurrence probability

To predict the likely distribution of *Aedes aegypti* mosquitoes, Kraemer et al.[7] generated high-resolution occurrence probability surfaces based on a species distribution modeling approach[11]. More specifically, a boosted regression tree model was applied using a comprehensive set of known occurrences (n = 19,930) of *Ae. aegypti* and a set of environmental predictors known to influence the distribution of the species[7]. Covariates included a temperature suitability index[17], contemporary mean and range maps of the Enhanced Vegetation Index and precipitation[34], and an urbanization index from the Global Rural Urban Mapping Project. We used a set of 100 spatial layers sampled from the posterior distribution estimated by Kraemer et al.[7].

Economic index

To account for socio-economic differences among populations residing in different regions, we used one-degree resolution gridded estimates of purchasing power parity (PPP) in U.S. Dollars from 2005 adjusted for inflation (G-Econ)[28]. When we encountered missing values, we imputed values in one of two ways. Grid cells in small island countries with data missing for the entire country were uniformly filled with population-adjusted PPP figures obtained from the U.S. CIA World Factbook[35]. Missing values in continental grid cells were imputed with the mean of the surrounding eight grid cell values. Once we obtained a complete PPP grid layer at one-degree resolution, we resampled the layer to a resolution of 5x5 km to match the resolution of gridded layers for human demography, temperature, and *Ae. aegypti* occurrence probability.

Seroprevalence estimates

To calibrate our model, we identified published estimates of seroprevlance that were relevant to the context of our study (Extended Data Table 1). Specifically, we sought estimates of seroprevalence to either Zika or chikungunya viruses in populations that were presumably naïve prior to an outbreak. Thus, we excluded some seroprevalence estimates that were obtained from endemic populations. We also excluded estimates from small islands—namely, Reunion and Grande Comore—for which it was clear that gridded temperature data were unrealistically low

due to steep elevational gradients and other features of island geography. Although the focus of our analysis was on Latin America and the Caribbean, we were not able to exclude locations on the basis of location given that only 2/13 came from the focal region. Appropriately however, a number of the seroprevalence estimates we obtained pertained specifically to pregnant women, although there did not appear to be differences in the seroprevalence of pregnant women and the population at large, at least in the context of a naïve population following an outbreak[36].

Calculation of derived quantities

Mosquito abundance

Occurrence probabilities can be translated into proxies for abundance provided that an assumption is made about how abundance is distributed as a random variable[37]. Assuming that mosquito abundance is distributed as a Poisson random variable, the probability that there is at least one mosquito present in a given location is $1 - \exp(-\lambda)$, where λ is the expected abundance of mosquitoes. Inverting this relationship, we obtained an estimate $\lambda = -\ln(1 - \text{occurrence probability})$ of expected mosquito abundance under the Poisson model and used this as a proxy for mosquito abundance in our calculations.

Mosquito-human ratios

The estimates of mosquito occurrence probability that we used incorporated a number of environmental variables[7]. They did not account for factors that modulate contact between mosquitoes and humans, however. Due in part to economic differences, factors such as air conditioning and piped water can drastically limit mosquito-human contact and virus transmission, even when mosquitoes are abundant[20]. We accounted for the effect of economic differences between locations by multiplying our proxy for mosquito abundance λ by a multiplication factor modeled with a function of the gross cell product economic index that we fitted to match our attack rate projections with published seroprevalence estimates. We fitted this function describing the multiplication factor using a shape constrained additive model (SCAM[38]), because relationships between mosquito-human ratios and economic indices should be monotonically decreasing, nonlinear, and without a predictable functional form. The response variable to which we fitted this relationship was a set of scalar multiples of expected mosquito abundances that would result in a perfect correspondence between attack rates derived from our model and published seroprevalence estimates.

Basic reproduction number R_0

We calculated the basic reproduction number R_0 according to its classic Ross-Macdonald formulation and as a function of temperature T,

$$R_0(T) = \frac{mbca^2 e^{-\mu(T)n(T)}}{\mu(T)r}, \qquad (1)$$

with adult mosquito mortality μ and extrinsic incubation period n specified as functions of temperature. Because temperature values were available for each location on a monthly basis, we computed monthly values of R_0 for each location and then used the mean of the highest six monthly values of R_0 as a singular estimate of R_0 for each location. This approach was broadly consistent with the way in which a temperature suitability index was used to inform mosquito occurrence probabilities by Kraemer et al.[7].

For mosquito mortality, we used the temperature- and age-dependent model of Brady et al.[39], to which we added an additional force of extrinsic mortality (0.025 d^{-1}) to match an overall daily mortality value of 0.115 estimated in a mark-release-recapture experiment carried out under temperatures ranging 20–34 °C [40]. We then computed the mean of the age- and temperature-dependent lifespan distribution as a function of temperature to inform $\mu(T)$. For the relationship between temperature and mean duration of the extrinsic incubation period, we used the temperature-dependent exponential rate estimated by Chan and Johansson[18]. The ratio of mosquitoes to humans, m, was quantified using a combination of occurrence probabilities and the gross cell product economic index, as described in the previous two sections. Parameters that did not depend on temperature were set at the following values according to published estimates for *Ae. aegypti* and dengue virus: mosquito-to-human transmission probability, $b = 0.4$ [41]; human-to-mosquito transmission probability times number of days of human infectiousness, $c / r = 3.5$ [42]; and mosquito biting rate, $a = 0.67$ [43]. Although there is uncertainty around these parameter values, any such uncertainty was effectively subsumed by fitting m to seroprevalence data given that bca^2/r entered R_0 as a constant.

Attack rates under a model-based formulation

Under a susceptible-infected-recovered (SIR) transmission model, there is a one-to-one relationship between R_0 and final epidemic size, which is equivalent to the attack rate over the course of an epidemic[13]. Intuitively, the final epidemic size is reached once herd immunity is sufficient to limit contacts between infectious and susceptible individuals to the extent necessary to reduce the pathogen's force of infection to zero. There is no explicit solution for final epidemic size as a function of model parameters, but it can be calculated numerically by obtaining an implicit solution of $S_\infty = e^{-R_0(1-S_\infty)}$ for S_∞, which is the proportion remaining susceptible after the epidemic has burned out[13]. Under the assumptions of the SIR model, the attack rate over the course of an epidemic is $AR = 1 - S_\infty$.

To apply this theoretical insight to Zika or other mosquito-borne pathogens, several limiting assumptions of the SIR model must first be reconciled. One such assumption is that individuals become infectious immediately upon becoming infected and remain infectious for an exponentially distributed period of time[44]; mosquito-borne pathogens such as Zika virus are instead characterized by a distinct lag between human and mosquito infection[45]. Despite this discrepancy between assumptions of the SIR model and the reality of many pathogen systems, mathematical analyses[46] have shown that final epidemic size is insensitive to details about the shape of the distribution that characterizes the time period between successive cases (i.e., the generation interval).

Another limiting assumption of the SIR model is that of homogeneous encounters between people and mosquitoes[44], which are understood to be extensive for mosquito-borne diseases[22]. Mathematical analyses[21] in this case show that a seemingly infinite complexity of relationships between R_0 and final epidemic size are possible in a heterogeneous system. As a general rule, however, final epidemic size in a system with contact heterogeneity and proportional mixing is expected to be strictly less than the final epidemic size in an otherwise equivalent system with homogeneous contacts[21]. How the ratio of these final epidemic sizes scales as a function of R_0 depends entirely on the details of a given system and has so far not been generalized mathematically.

To capture the potentially very strong effects of heterogeneity in reducing final epidemic size in populations subject to Zika epidemics, we scaled final epidemic size by substituting R_0^{α} for R_0 in the SIR-based final epidemic size formula given some constant $\alpha \in (0,1]$. Although there is no theoretical justification for this or any other choice of how to scale R_0 and AR in the presence of contact heterogeneity, the choice we made has the following desirable properties: (1) it implies that $AR \to 1$ as $R_0 \to \infty$; (2) it leads to the function $AR(R_0)$ having a more gradual slope and thereby allowing for intermediate attack rates to be more common than they would be otherwise; and (3) it preserves the property that $AR = 0$ for $R_0 < 1$. At the same time, this and possible alternative formulations are limited by a general lack of understanding about the relationship between R_0 and AR in heterogeneous systems, relationships that may furthermore be heterogeneous themselves across different areas[47].

To estimate α, we performed the following procedure for candidate values of α between 0.01 and 1 in increments of 0.01: (1) calculate R_0 according to eqn. 1 and assuming $m = 1$ for each of the 13 sites from which seroprevalence estimates were derived; (2) use those R_0 values to calculate AR values for each of those sites based on the classic SIR formulation; (3) calculate what multiplication factor of R_0 would be necessary for AR to match the empirical seroprevalence estimate; (4) fit a SCAM model of the economic index to the multiplication factors; and (5) use the fitted SCAM values to recalculate R_0 and then AR for each site. Next, we calculated the sum of squares between the final predicted AR values associated with each α and the empirical seroprevalence estimates, and we then selected the value of α that minimized the sum of squares. Extended Data Figure 1 illustrates this process given mean estimates of *Ae. aegypti* occurrence probabilities, $\mu(T)$, and $n(T)$.

<u>Attack rates under a statistical formulation</u>

As an alternative to our model-based characterization of attack rates, we also considered a purely statistical approach that modeled probit-transformed seroprevalence observations as functions of averaged monthly temperatures, *Ae. aegypti* occurrence probabilities, and the economic index. We considered all combinations of linear, quadratic, and pairwise interaction terms of these variables, comparing them on the basis of Akaike Information Criterion using the lm and step functions in R[48]. Although additional functional forms would have been of interest, this suite of models was as complex as the limited set of 13 seroprevalence observations would support.

<u>Quantifying uncertainty around attack rate projections</u>

To quantify uncertainty associated with our projections, we generated 1,000 Monte Carlo samples from the uncertainty distributions of each model parameter. For $\mu(T)$ and $n(T)$, we took random draws of their parameters consistent with published descriptions of uncertainty in the parameters of these functions from their original sources[17,18]. For *Ae. aegypti* occurrence probabilities, we drew randomly with replacement from 100 sample layers from the posterior distribution[7]. For the relationship involving the economic index and the R_0 scaling factor α, we used best-fit SCAM models and α values corresponding to each set of random draws of the parameters of $\mu(T)$, $n(T)$, and the *Ae. aegypti* layers. For each of the 1,000 Monte Carlo samples of the statistical model, we performed resampling with replacement among the 13 seroprevalence values, performed the same model fitting and model selection procedure described in the

previous section, and took a multivariate normal random sample of the parameter values of the best-fit model based on the model's best-fit parameters and variance-covariance matrix.

Projecting attack rates and numbers of infections

To obtain estimates of numbers of infections in total and among childbearing women for the model-based and statistical approaches, we multiplied their respective attack rate projections applied to 5x5 km grids across Latin America and the Caribbean by human demographic layers for total population and births in 2015. For both the model-based and statistical approaches, we performed these calculations and summed at the country level once for each of the 1,000 Monte Carlo samples that we produced. High-resolution spatial projections of attack rates and numbers of infected childbearing women under the model-based approach are presented in Extended Data Figs. 2-10. Most projections based on the statistical approach resulted in attack rates of 100% in nearly all locations throughout Latin America and the Caribbean.

REFERENCES

1. Mlakar, J., et al. Zika virus associated with microcephaly. *New England Journal of Medicine* **374**, 951-958 (2016).
2. Brasil, P., et al. Zika virus infection in pregnant women in Rio de Janeiro—preliminary report. *New England Journal of Medicine* doi:10.1056/NEJMoa1602412 (2016).
3. Messina, J.P., et al. Mapping global environmental suitability for Zika virus. *eLife* doi:10.7554/eLife.15272 (2016).
4. Cauchemez, S., et al. Association between Zika virus infection and microcephaly in French Polynesia, 2013-2015: a retrospective study. *Lancet* doi:10.1016/S0140-6736(16)00651-6 (2016).
5. Johansson, M.A., et al. Risk estimates for microcephaly related to Zika virus infection – from French Polynesia to Bahia, Brazil. *bioRxiv* doi:10.1101/051060.
6. World Health Organization. WHO statement on the first meeting of the International Health Regulations (2005) (IHR 2005) Emergency Committee on Zika virus and observed increase in neurological disorders and neonatal malformations. 1 February 2016, http://www.who.int/mediacentre/news/statements/2016/1st-emergency-committee-zika/en/.
7. Kraemer, M. U. G., et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife* e08347 (2015).
8. Pan American Health Organization. Zika – Epidemiological Update. April 28, 2016.
9. Lucey, D. R. & Gostin, L. O. The Emerging Zika Pandemic: Enhancing Preparedness. *Journal of the American Medical Association* (2016).
10. Editorial. The next steps on Zika. *Nature* **530**, 5 (2016).
11. Elith, J. & Leathwick, J. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**, 677 (2009).
12. Kermack, W. & McKendrick, A. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A* **115**, 700-721 (1927).
13. Keeling, M. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. Princeton (2007).
14. Sorichetta, A., et al. High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific Data* **2**, 150045 (2015).
15. Sergon, K., et al. Seroprevalence of chikungunya virus (CHIKV) infection on Lamu Island, Kenya, October 2004. *American Journal of Tropical Medicine and Hygiene* **78**, 333-337 (2008).
16. Smith, D. L., et al. Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLOS Pathogens* **8**, e1002588 (2012).
17. Brady, O. J., et al. Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasites and Vectors* **7**, 338 (2014).

18. Chan, M. & Johansson, M. A. The incubation periods of dengue viruses. *PLOS ONE* **7**, e50972 (2012).
19. Funk, S., et al. Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus. *bioRxiv* doi:10.1101/043265 (2016).
20. Reiter, P., et al. Texas lifestyle limits transmission of dengue virus. *Emerging Infectious Diseases* **9**, 86-89 (2003).
21. Andreasen, V. The final size of an epidemic and its relation to the basic reproduction number. *Bulletin of Mathematical Biology* **73**, 2305-2321 (2011).
22. Liebman, K. A., et al. Determinants of heterogeneous blood feeding patterns by *Aedes aegypti* in Iquitos, Peru. *PLOS Neglected Tropical Diseases* **8**, e2702 (2014).
23. S. Bhatt, et al. The global distribution and burden of dengue. *Nature* **496**, 504-507 (2013).
24. Feldstein, L., Brownstein, J., Brady, O.J., Hay, S. I. & M. A. Johansson. Dengue on islands: a Bayesian approach to understanding the global ecology of dengue viruses. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **109**, 303-312 (2015).
25. UNFPA. The State of the World's Midwifery 2014. (2014).
26. Tatem, A.J., et al. Millenium development health metrics: where do Africa's children and women of childbearing age live? *Population Health Metrics* **11**, 11 (2013).
27. Hijmans, R. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978 (2005).
28. Nordhaus, W. Geography and macroeconomics: new data and new findings. *Proceedings of the National Academy of Sciences* 103, 3510-3517 (2006).
29. Taylor, B.P., et al. Stochasticity and the limits to confidence when estimating R_0 of Ebola and other emerging infectious diseases. arXiv:1601.06829 (2016).
30. Metcalf, C.J.E., et al. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet* doi:10.1016/S0140-6736(16)30164-7 (2016).
31. Tatem, A. J., et al. Mapping for maternal and newborn health: the distributions of women of childbearing age, pregnancies and births. *International Journal of Health Geographics* **13**, 2 (2013).
32. Stevens, F.R., et al. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLOS ONE* **10**, e0107042 (2015).
33. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2015 Revision. (2015).
34. Weiss, D. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing* **98**, 106-118 (2014).
35. U. S. Central Intelligence Agency. The World Factbook. 5 January 2016. https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html.
36. Gerardin, P. Estimating chikungunya prevalence in Reunion Island outbreak by serosurveys: two methods for two critical times of the epidemic. *BMC Infectious Diseases* **8**, 99 (2008).
37. Wright, D. Correlations between incidence and abundance are expected by chance. *Journal of Biogeography* **18**, 463-466 (1991).
38. Pya, N. & Wood, S. Shape constrained additive models. *Statistics and Computing* **25**, 543-559 (2015).
39. Brady, O. J., et al. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasites and Vectors* **6**, 351-362 (2013).
41. Muir, L. & Kay, B. *Aedes aegypti* survival and dispersal estimated by mark-release-recapture in northern Australia. *American Journal of Tropical Medicine and Hygiene* **58**, 277-282 (1998).
42. Nishiura, H., & Halstead, S. Natural history of dengue virus (DENV)-1 and DENV-4 infections: reanalysis of classic studies. *Journal of Infectious Diseases* **195**, 1007-1013 (2007).
43. Scott, T.W., et al. Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: blood feeding frequency. *Journal of Medical Entomology* **37**, 89-101 (2000).
44. Reiner, R. C., et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. *Journal of the Royal Society Interface* **10**, 20120921 (2013).

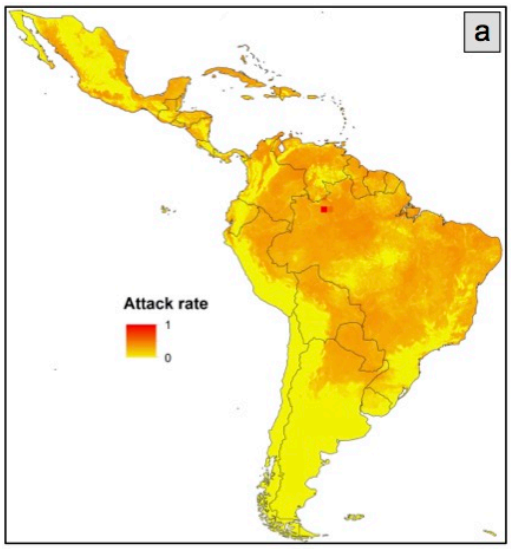
45. Lessler, J., et al. Times to key events in the course of Zika infection and their implications for surveillance: a systematic review and pooled analysis. *bioRxiv* doi:10.1101/041913.
46. Ma, J. & Earn, D.J. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology* **68**, 679-702 (2006).
47. Kraemer, M.U.G., et al. Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *Journal of the Royal Society Interface* **12**, 20150468 (2015).
48. R Core Team. R: A language for statistical computing. R Foundation for Statistical Computing. Vienna, Austria (2014).
49. Duffy, M., et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. *New England Journal of Medicine* **360**, 2536-2543 (2009).
50. Schwarz, N., et al. Seroprevalence of antibodies against chikungunya, dengue, and Rift Valley fever viruses after febrile illness outbreak, Madagascar. *Emerging Infectious Diseases* **18**, 1780-1786 (2012).
51. Sissoko, D. Outbreak of chikungunya fever in Mayotte, Comoros archipelago, 2005-2006. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **102**, 780-786 (2008).
52. Dwibedi, B., et al. Rapid spread of chikungunya virus infection in Orissa, India. *Indian Journal of Medical Research* **133**, 316-321 (2011).
53. Gay, N., et al. Seroprevalence and asymptomatic rates of Asian lineage chikungunya virus infection on Saint Martin, Caribbean. *American Journal of Tropical Medicine and Hygiene* **94**, 393-396 (2015).
54. Moro, M., et al. Chikungunya virus in north-eastern Italy: a seroprevalence study. *American Journal of Tropical Medicine and Hygiene* **82**, 508-511 (2010).
55. Balmaseda, A., et al. Clinical attack rate of chikungunya in a cohort of Nicaraguan children. *American Journal of Tropical Medicine and Hygiene* **94**, 397-399 (2016).

ACKNOWLEDGEMENTS

We thank J Ashander, CM Barker, MA Johansson, RC Reiner, ST Stoddard, and members of the Perkins Lab for useful discussions. TAP, ASS, and AJT are supported by funding from NSF (DEB 1641130). TAP is supported by funding from NIH/NIAID (1P01AI098670-01A1) and the Bill & Melinda Gates Foundation (OPP1110495). AJT is supported by funding from NIH/NIAID (U19AI089674), the BMGF (OPP1106427, 1032350), NORAD, and a Wellcome Trust Sustaining Health Grant (106866/Z/15/Z). CWR is supported by funding through the University of Southampton's Economic and Social Research Council's Doctoral Training Centre. AJT and CWR acknowledge the support of the WorldPop (www.worldpop.org) and Flowminder Foundation (www.flowminder.org) teams in demographic dataset production, and TAP and ASS acknowledge support from the Notre Dame Center for Research Computing.

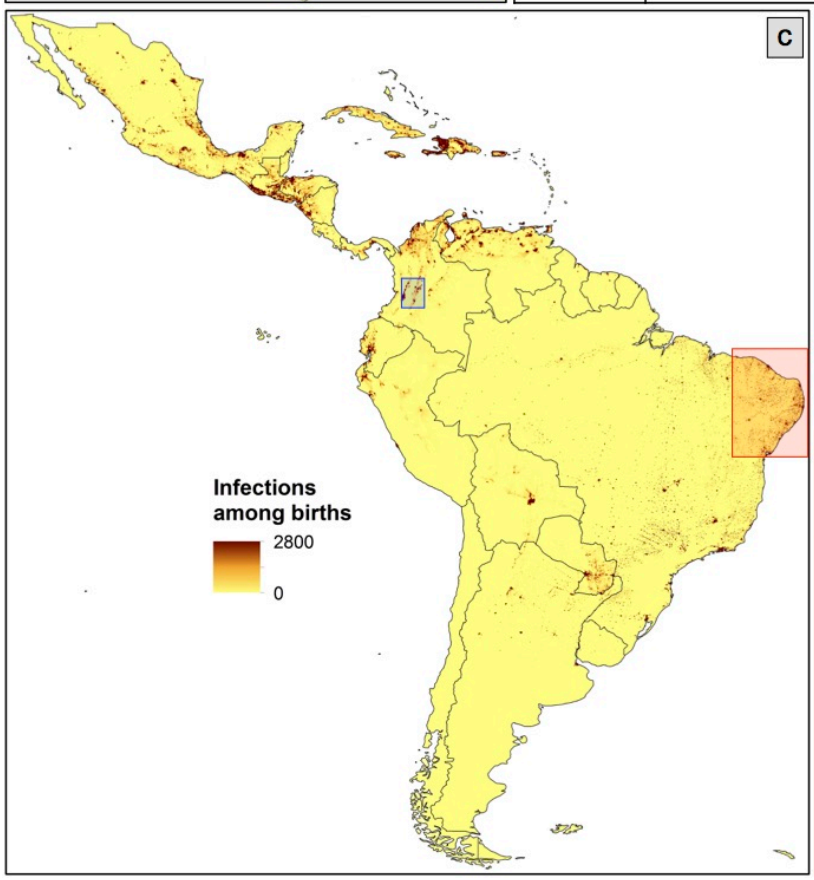
FIGURES

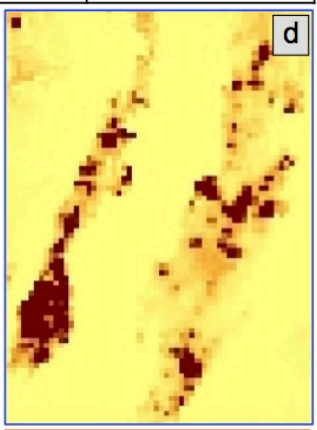
Figure 1. Spatial projections of location-specific epidemic attack rates (a) that combine with demographic projections[14] to yield projections of total numbers of Zika infections in all people and in childbearing women, which can either be summarized by country (b) or on a map (c). More detailed spatial projections of infections among childbearing women are shown for two areas: Cali, Colombia (d) and Recife, Brazil (e).



b

Projected Zika virus infections in first-wave epidemic		
Country	**Total ($x10^6$)**	**Births ($x10^3$)**
Brazil	37.4 (30.6-49.2)	579 (473-755)
Mexico	14.9 (12.8-17.5)	263 (226-310)
Venezuela	7.4 (6.6-8.2)	144 (130-162)
Colombia	6.7 (6.0-8.1)	124 (110-153)
Cuba	3.7 (3.3-4.1)	33 (30-37)
Haiti	2.9 (2.5-4.5)	73 (63-121)
Argentina	2.7 (1.6-6.6)	44 (27-108)
Dom. Rep.	2.6 (2.4-2.8)	61 (56-68)
Others	15.6 (13.3-20.8)	336 (285-453)
Total	**93.4 (81.6-117.1)**	**1,650 (1,449-2,062)**





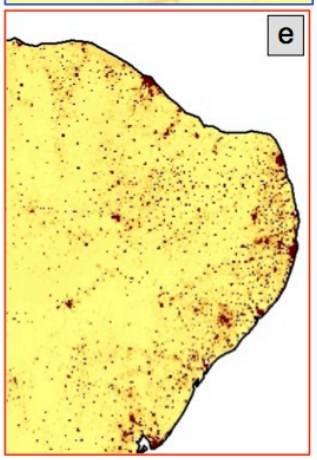


Figure 2. Monte Carlo samples from the uncertainty distributions surrounding each of the key drivers in the model (a-e) and uncertainty distributions for projected numbers of infections among all individuals (f-j) and among childbearing women (k-o) in different areas. All panels reflect the full range of uncertainty considered in 1,000 Monte Carlo samples. Panel a shows posterior distributions of mosquito occurrence probabilities for two example 5x5 km grid cells.

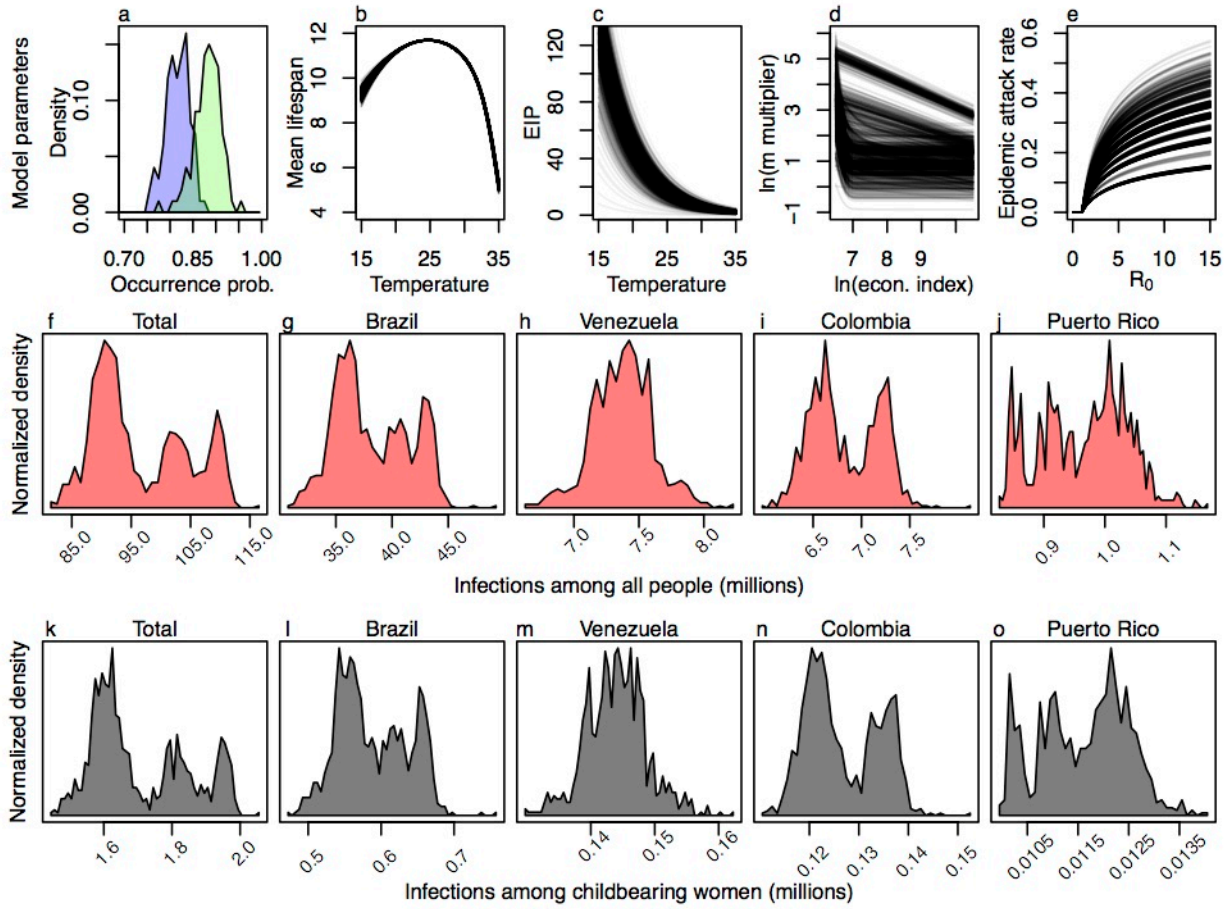


Figure 3. Projected numbers of infections among all individuals (a-e) and among childbearing women (f-j) for 1,000 Monte Carlo samples from the uncertainty distribution around parameters of the statistical model that was fitted to seroprevalence estimates from Extended Data Table 1.

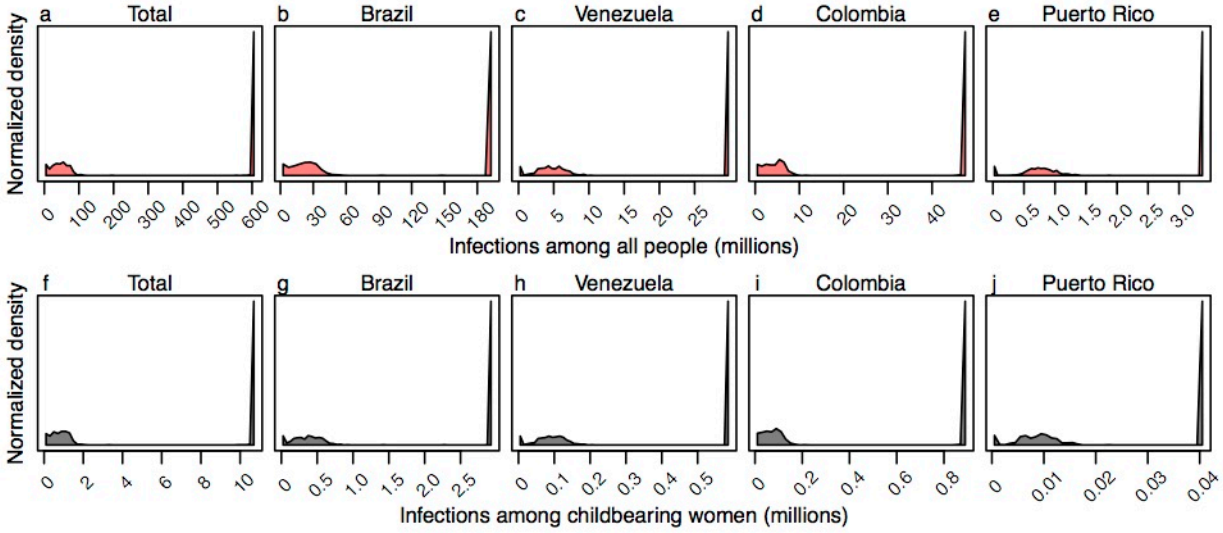
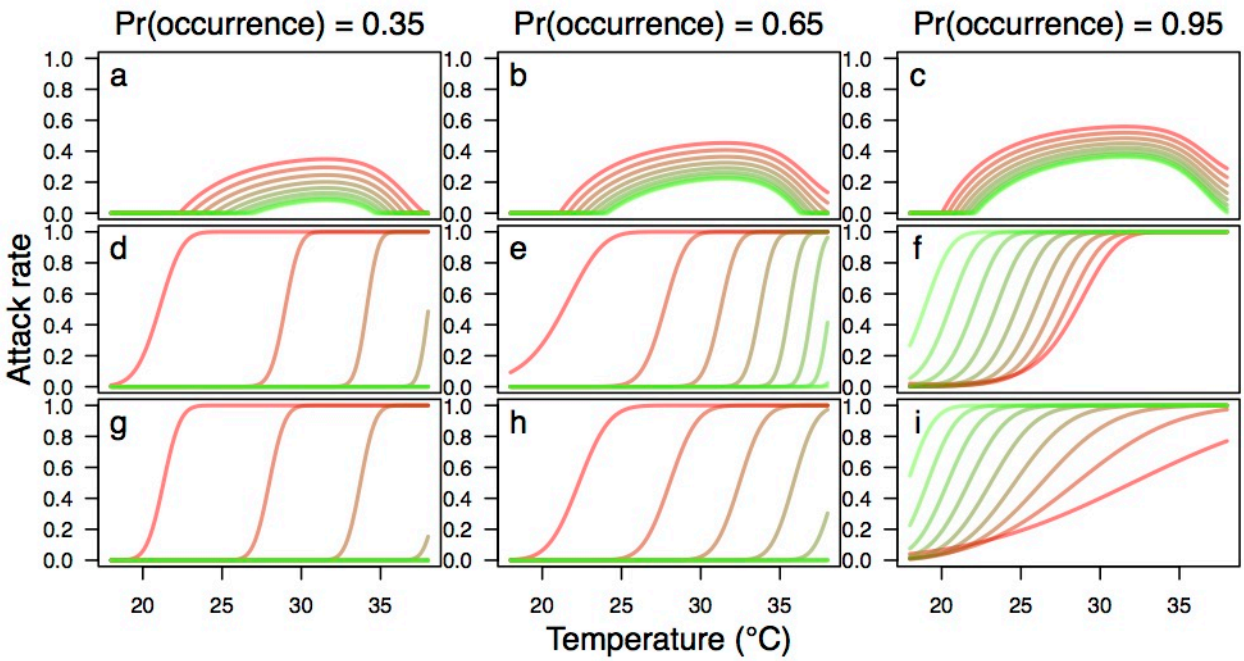


Figure 4. Relationships between temperature (x-axis), economic index (red = low, green = high), mosquito occurrence probabilities (columns), and projected epidemic attack rates (y-axis). These relationships are shown for the model-based approach (a-c), the full statistical approach (d-f), and the version of the statistical approach chosen by stepwise model selection (g-i).



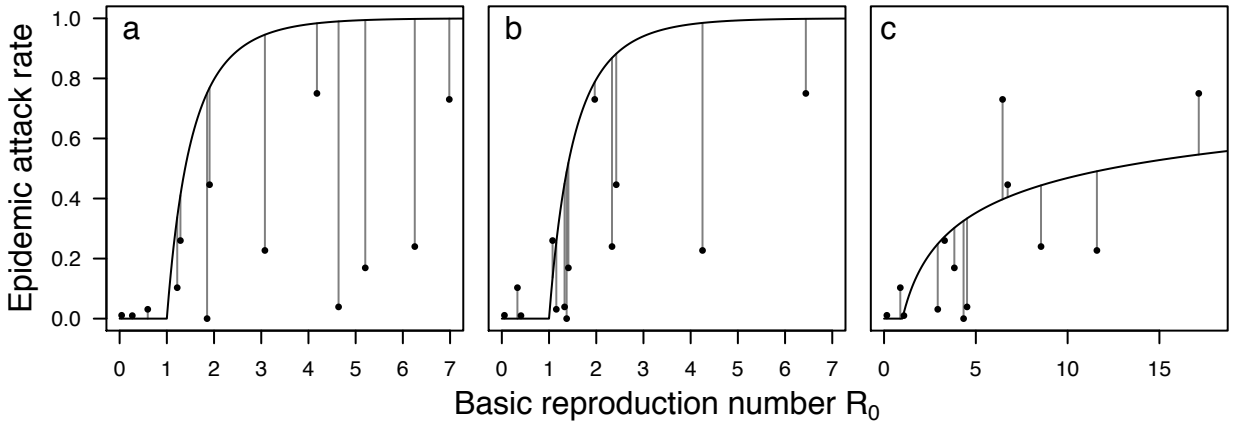
EXTENDED DATA

Extended Data Table 1. Seroprevalence estimates for Zika and chikungunya viruses in the context of recent oubtreaks in a previously susceptible population.

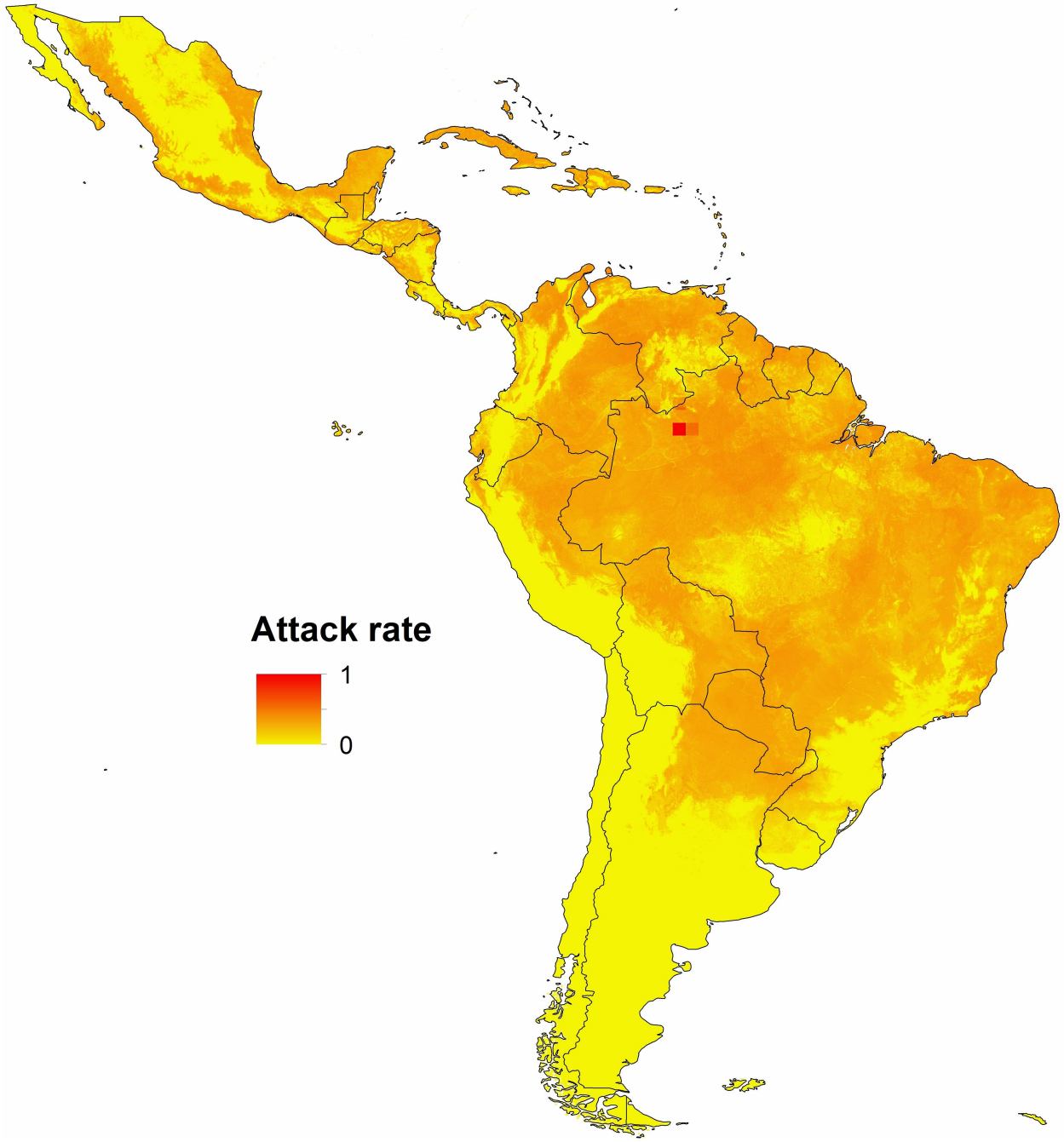
Seroprevalence	Virus	Location	Reference
0.75	CHIKV	Lamu Island, Kenya	15
0.73	ZIKV	Yap Island, Federated States of Micronesia	49
0.446	CHIKV	Mananjary, Madagascar	50
0.26	CHIKV	Mayotte Island, Union of the Comoros	51
0.24	CHIKV	Orissa, India	52
0.227	CHIKV	Manakara, Madagasar	50
0.169	CHIKV	Saint Martin	53
0.103	CHIKV	Emilia-Romagna, Italy	54
0.039*	CHIKV	Managua, Nicaragua	55
0.031	CHIKV	Moramanga, Madagascar	50
0.011	CHIKV	Ambositra, Madagascar	50
0.01	CHIKV	Ifanadiana, Madagascar	50
0.0	CHIKV	Tsiroanomandidy, Madagascar	50

* This estimate was based on a clinical attack rate of 0.029 and an assumed reporting rate of 0.75.

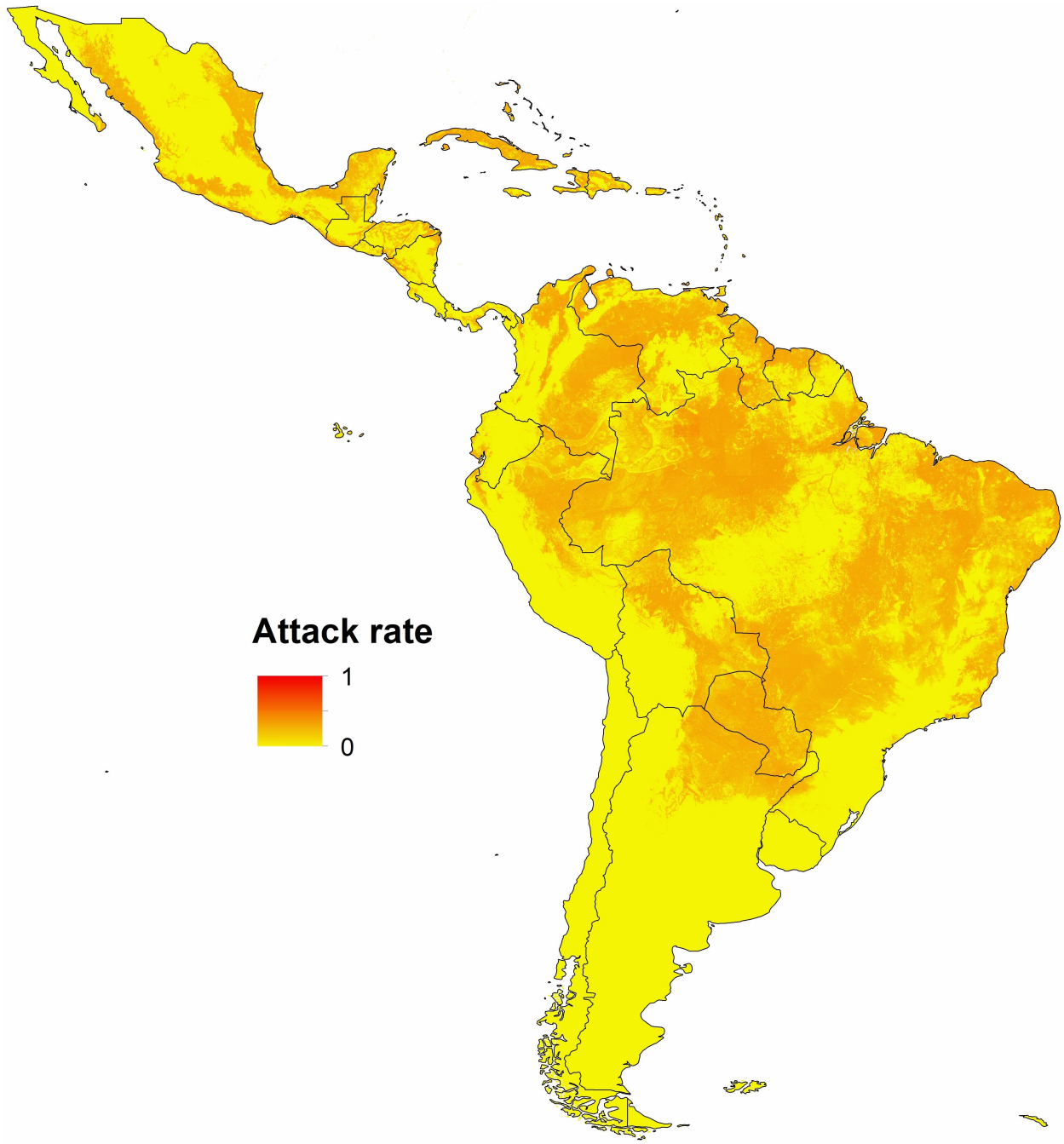
Extended Data Figure 1. Relationships between the basic reproduction number R_0 (x-axis) and epidemic attack rates (y-axis) under different model assumptions (a-c). In each panel, the curve shows this relationship under a given set of model assumptions and the points represent the 13 seroprevalence estimates (ED Table 1) mapped on to the R_0-axis based on conditions at the sites where those data were collected. (a) R_0 calculations according to eqn. 1 under the assumption that $m = 1$ and with parameter $\alpha = 1$ determining the shape of the curve. (b) R_0 calculations according to eqn. 1 under the assumption that m is determined by a SCAM model of the economic index and with parameter $\alpha = 1$. (c) Same as b but with with parameter $\alpha = 0.13$ fitted by least squares and a separate SCAM model fitted conditional on α.



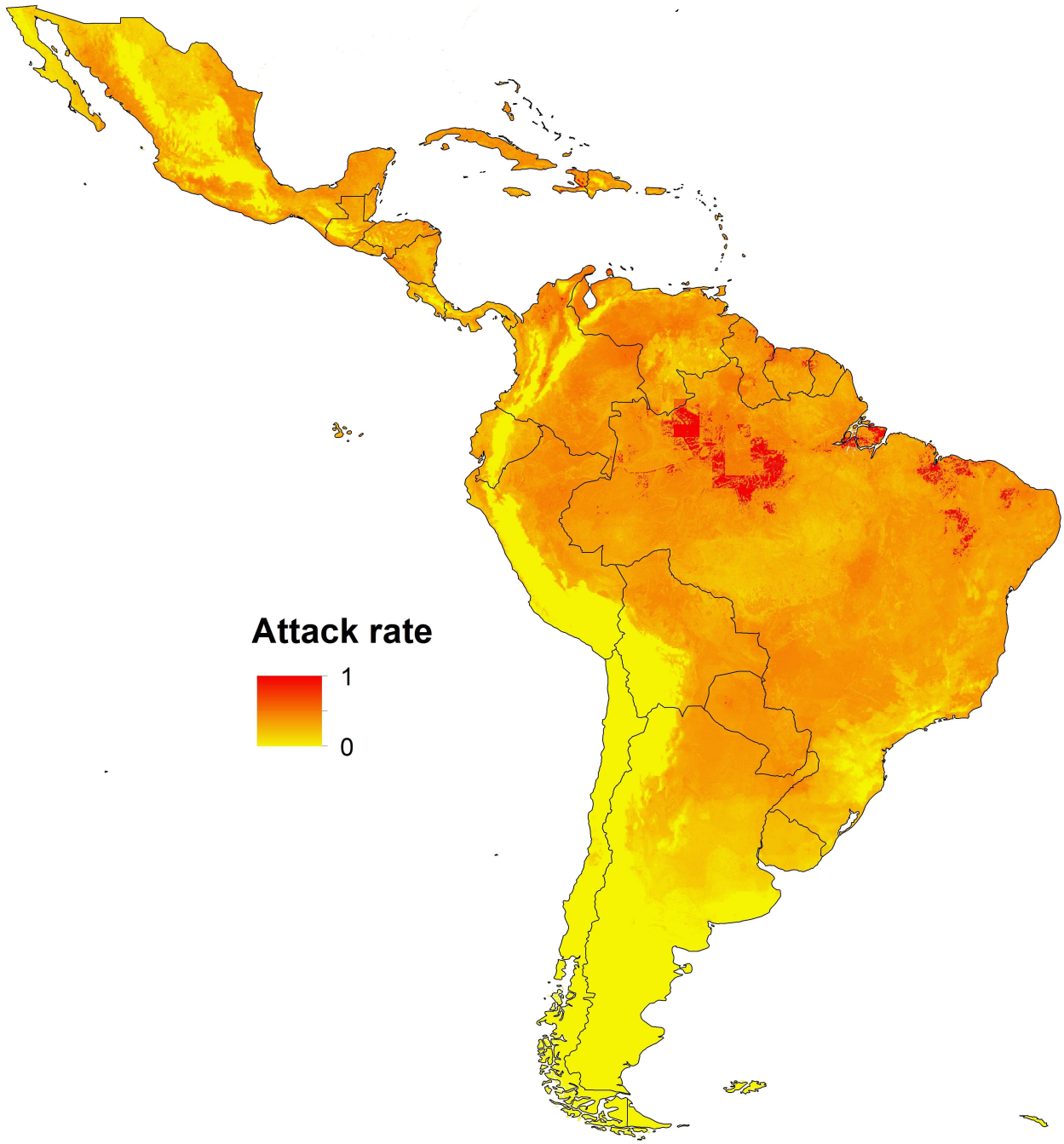
Extended Data Figure 2. Gridded spatial projections of median epidemic attack rates at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the median epidemic attack rate for that grid cell from the distribution of 1,000 Monte Carlo samples.



Extended Data Figure 3. Gridded spatial projections of minimum epidemic attack rates at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the minimum epidemic attack rate for that grid cell from the distribution of 1,000 Monte Carlo samples.

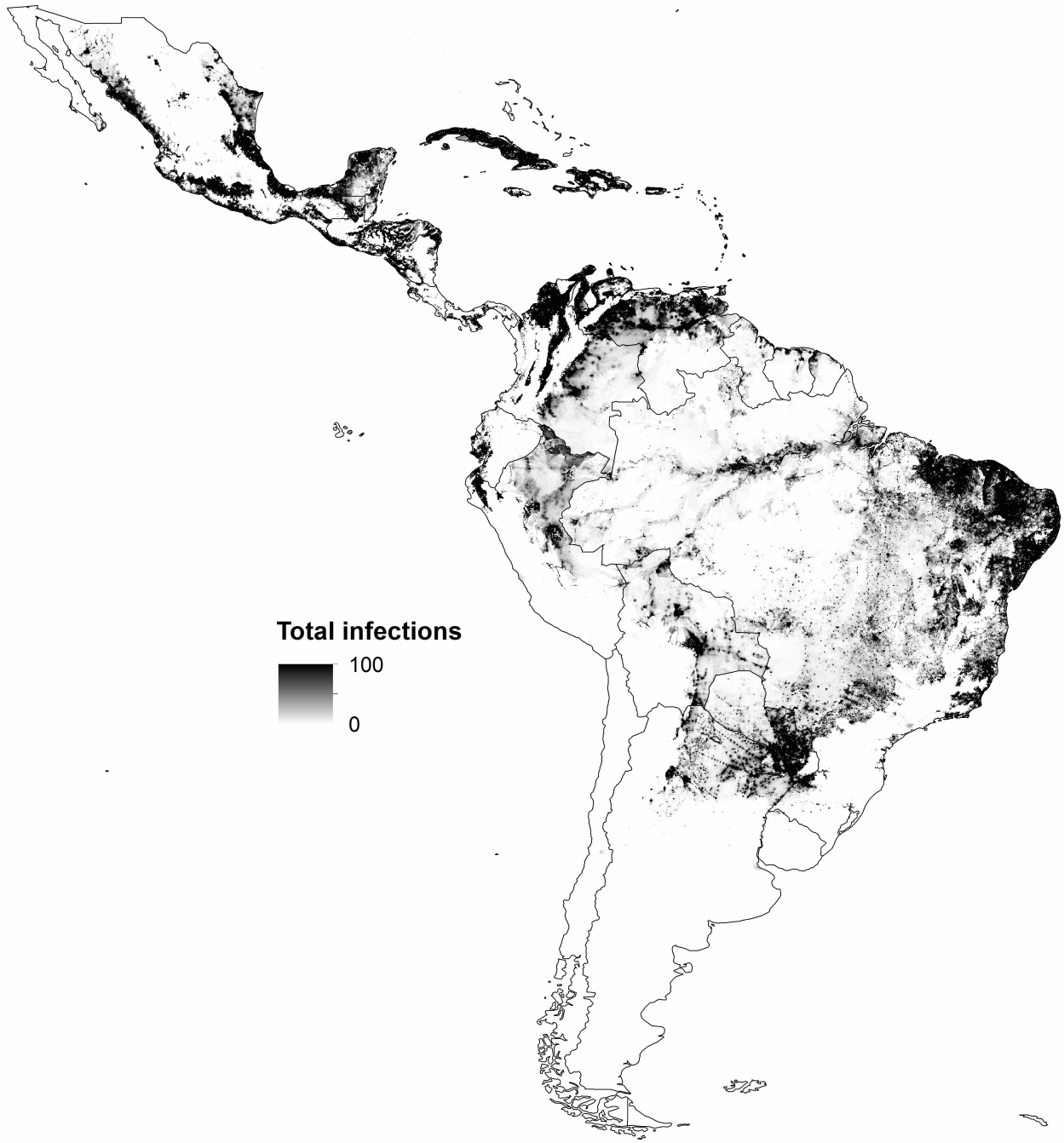


Extended Data Figure 4. Gridded spatial projections of maximum epidemic attack rates at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the maximum epidemic attack rate for that grid cell from the distribution of 1,000 Monte Carlo samples.



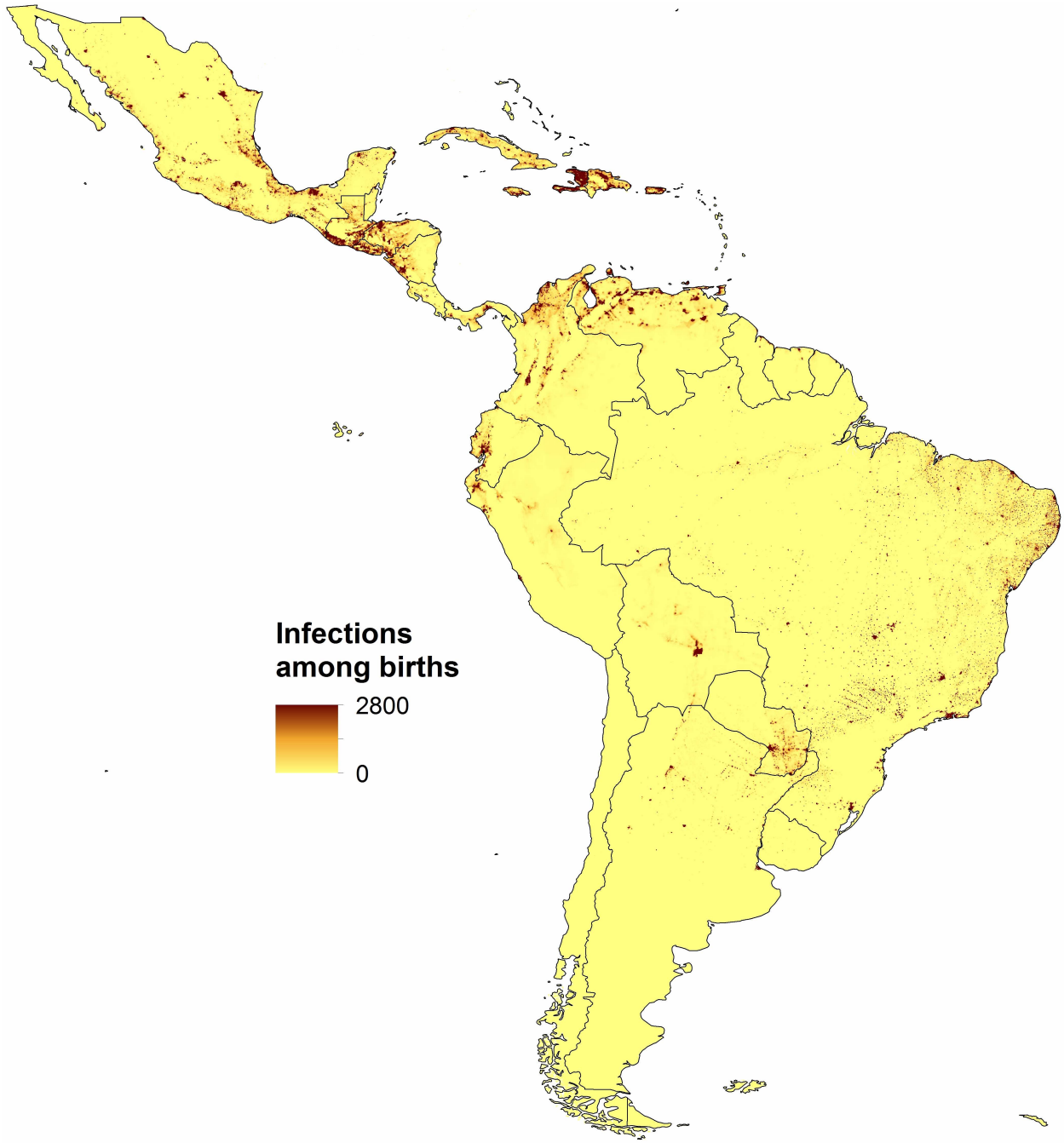
Extended Data Figure 5. Gridded spatial projections of median infections among the total population at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the median number of infections for that grid cell from the distribution of 1,000 Monte Carlo samples.

Extended Data Figure 6. Gridded spatial projections of minimum infections among the total population at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the minimum number of infections for that grid cell from the distribution of 1,000 Monte Carlo samples.

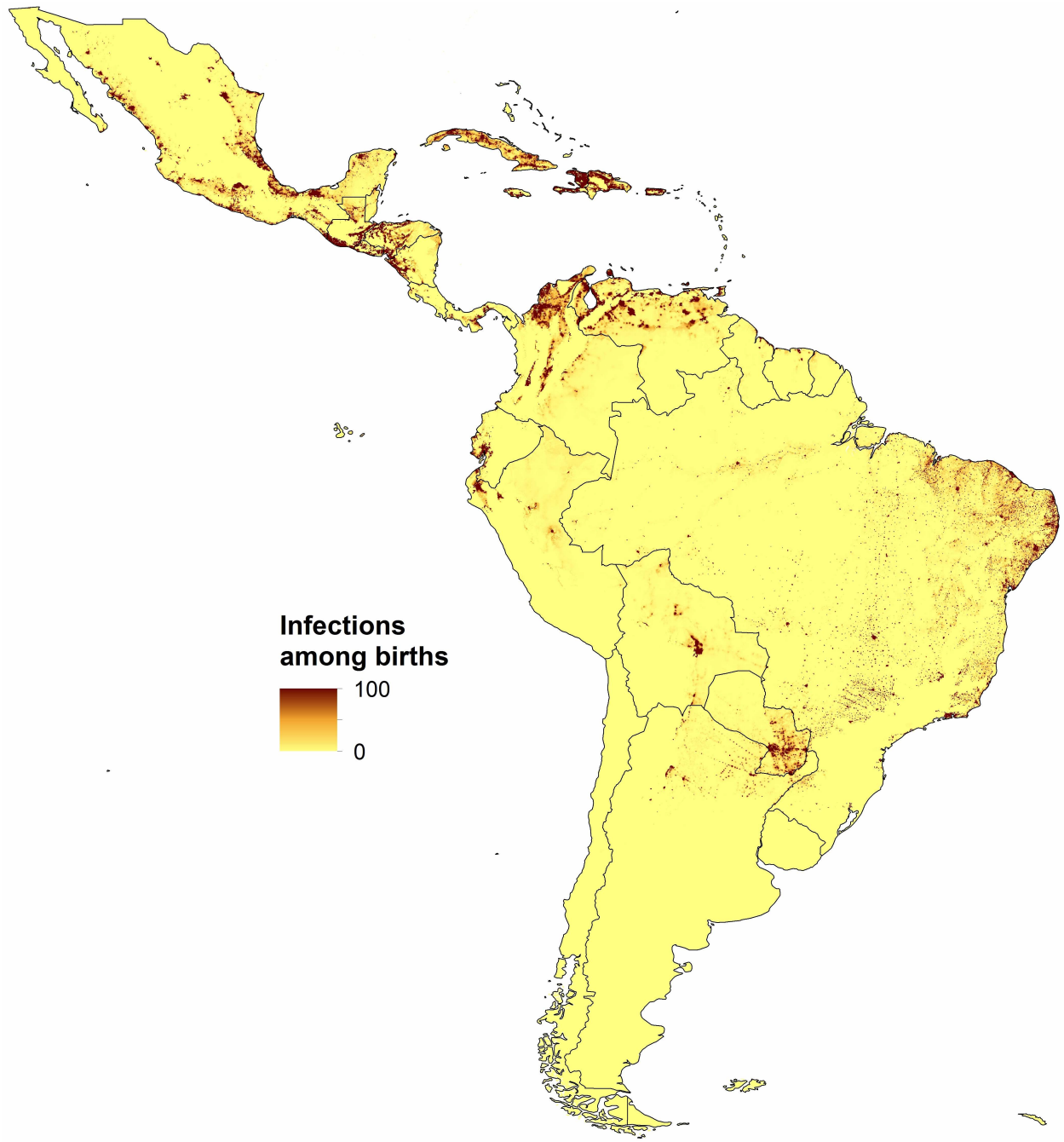
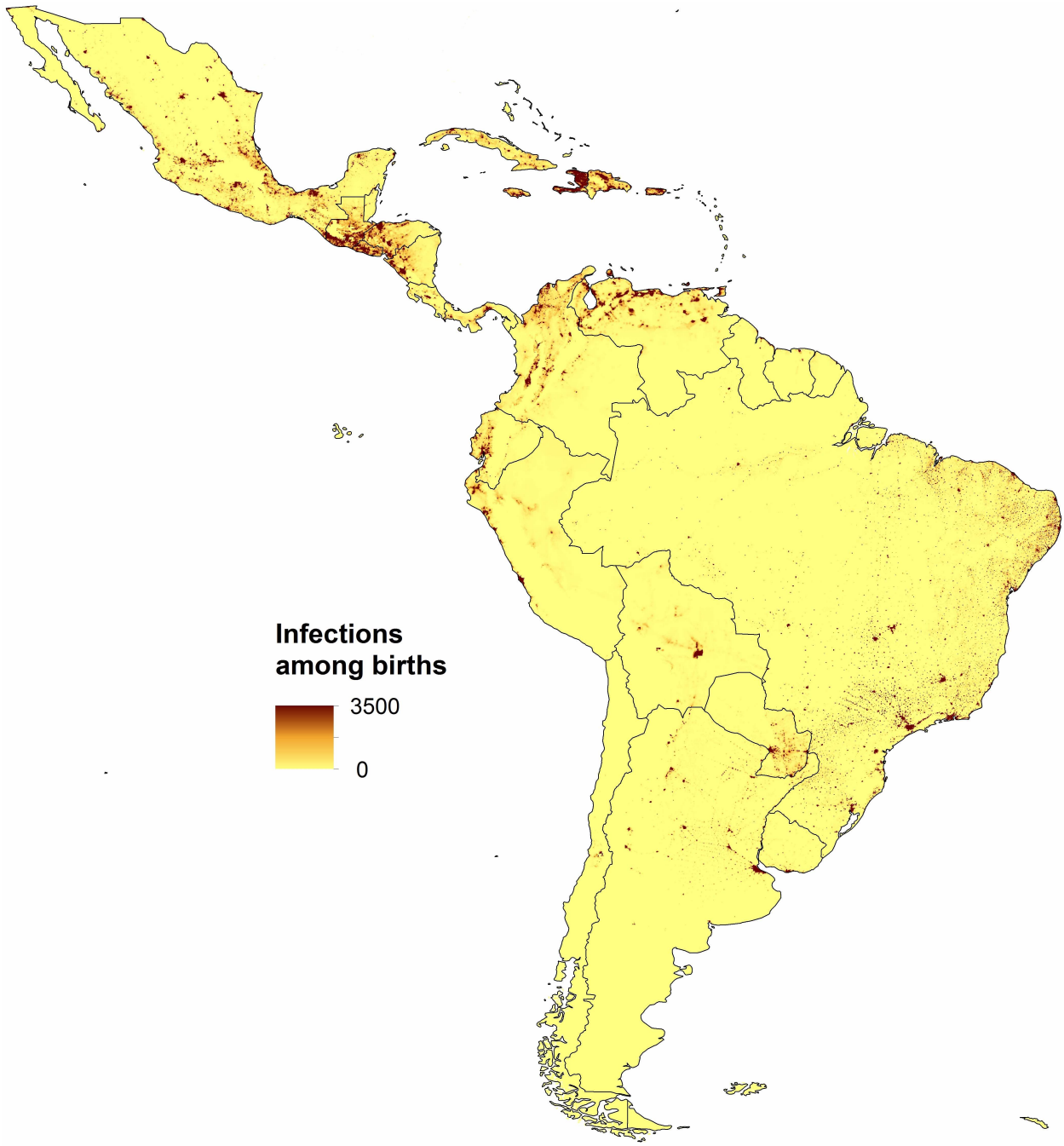


Extended Data Figure 7. Gridded spatial projections of maximum infections among the total population at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the maximum number of infections for that grid cell from the distribution of 1,000 Monte Carlo samples.

Extended Data Figure 8. Gridded spatial projections of median infections among childbearing women at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the median number of infections for that grid cell from the distribution of 1,000 Monte Carlo samples.



Extended Data Figure 9. Gridded spatial projections of minimum infections among childbearing women at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the minimum number of infections for that grid cell from the distribution of 1,000 Monte Carlo samples.



Extended Data Figure 10. Gridded spatial projections of maximum infections among childbearing women at 5x5 km resolution in Latin America and the Caribbean. Each grid cell is shaded according to the maximum number of infections for that grid cell from the distribution of 1,000 Monte Carlo samples.