Supplement: Real-time Zika risk assessment in the United States

Lauren A Castro [1*], Spencer J Fox[1*@], Xi Chen[2], Kai Liu[3], Steve Bellan[4], Nedialko B Dimitrov[2], Alison P Galvani[5,6], Lauren Ancel Meyers[1,7],

1 Section of Integrative Biology, The University of Texas at Austin, Austin, TX, USA
2 Graduate Program in Operations Research Industrial Engineering, The University of Texas at Austin, Austin, TX, USA
3 Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin,TX, USA
4 Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, TX, USA
5 Center for Infectious Disease Modeling and Analysis, Yale School of Public Health, New Haven, CT, USA
6 Department of Ecology and Evolution, Yale University, New Haven, CT, USA
7 The Santa Fe Institute, Santa Fe, NM, USA

*contributed equally to this manuscript
@ corresponding author: spncrfx@gmail.com

Methodological Overview

This manuscript introduces a framework for interpreting Zika Virus (ZIKV) surveillance data in the face of uncertainty regarding importation, transmission, and reporting rates. In this supplement, we provide additional methodological details (Sections 1-3), Texas county risk assessments (Section 3), and parameter sensitive analyses (Section 4). Sections 1-3 detail each of the three major steps of our methodology: (1) estimating county-specific ZIKV transmission and importation risks, (2) simulating ZIKV outbreaks using a probabilistic branching process ZIKV transmission model, and (3) analyzing simulations to estimate risk and provide analytic tools for interpreting surveillance data (Fig 1).

1 Texas County Risk Assessment

1.1 County Importation Rate Estimation

We assumed that ZIKV outbreaks originate with the arrival (importation) of an infectious individual and built a model to estimate county-level ZIKV importation rates across Texas. The model consists of two components, estimated separately and then multiplied to obtain county importation rates: (1) a statewide ZIKV importation rate and (2) county-specific probabilities of receiving the next Texas importation. The main text describes our estimation of statewide ZIKV importation rates; here, we detail our methods for estimating county-level probabilities. County-to-county variation in importation rates may be influenced by a number of different environmental, socioeconomic, and behavioral factors. To build a predictive model consisting of a tractable number of informative

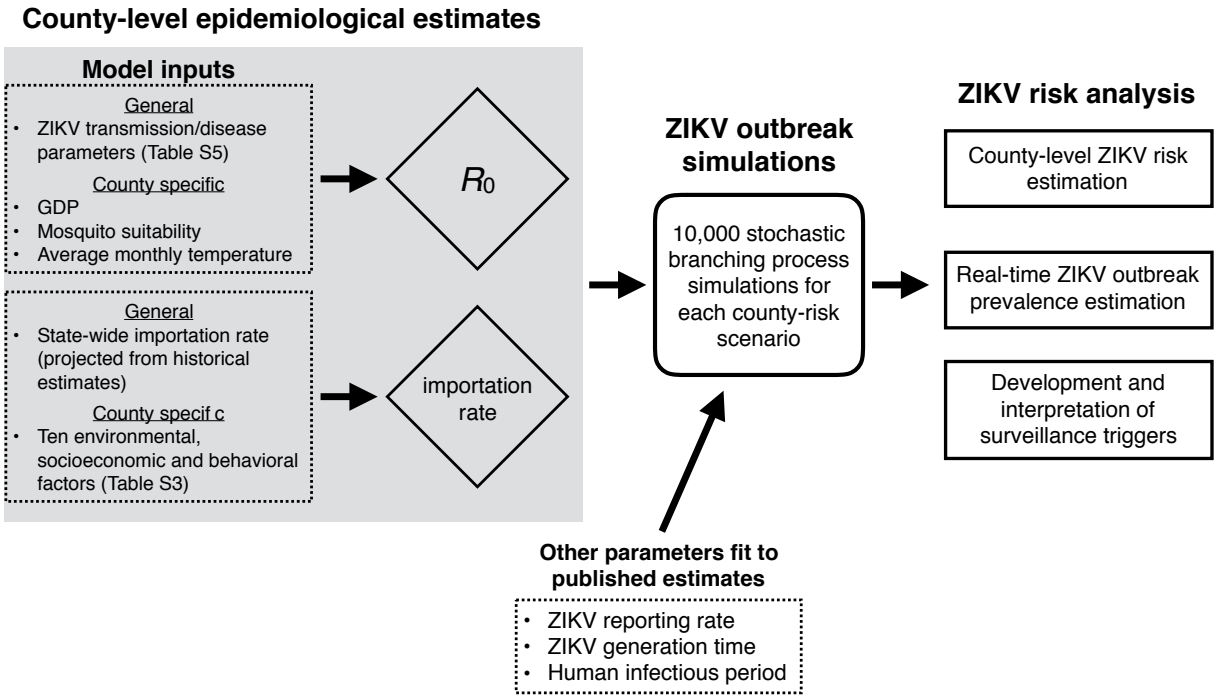

Figure 1: **ZIKV Risk Assessment Framework.** The method consists of three steps. First, we use data-derived models to estimate county-level ZIKV introduction rates and ZIKV transmission rates. Each estimate is based on a combination of general and county-specific factors. Second, for every county-risk combination, we simulate 10,000 ZIKV outbreaks using a stochastic branching process ZIKV transmission model parameterized by the county-level importation and transmission rate estimates along with several other recently published disease progression estimates. The simulations track the numbers of autochthonous and imported cases (unreported and reported) and, based on the total size and maximum daily prevalence, classifies each outbreak as self-limiting or epidemic. Third, we analyze the simulations to determine (1) robust relationships between the number of reported cases in a county and the current and future ZIKV threat and (2) surveillance triggers (number of reported cases) indicative of imminent epidemic expansion.

predictors, we fit a maximum entropy model consisting of dozens of possible predictors to historical arbovirus importation data, and then pruned the model through two rounds of variable selection.

1.1.1 Historical Arbovirus Data

We analyzed ten years (2002-2012) of Dengue (DENV) importation data (number of importations in a county) provided by the Texas Department of State Health Services (DSHS), two years of CHIKV importation data scraped from online DSHS reports, and the 30 recent ZIKV importations into Texas counties from January 2, 2016 to April 1, 2016 [1, 2]. Given the geographic and biological overlap between ZIKV, DENV, and CHIKV, we use historical DENV and CHIKV importation data to supplement ZIKV importations. We believe this decision is justified, as an earlier version of our importation risk model fit only to DENV and CHIKV data predicted the county ZIKV importation distribution well [3].

Currently, DENV, CHIKV, and ZIKV importation patterns differ most noticeably along the Texas-Mexico border. Endemic DENV transmission and sporadic CHIKV outbreaks in Mexico historically have spilled over into neighboring Texas counties. We included DENV and CHIKV importation data in the model fitting so as to consider potential future importations pressure from Mexico, as ZIKV cases continue to increase in Mexico.

1.1.2 Sociological Predictor Data

As predictors for county-level ZIKV importation risk, we included socioeconomic and demographic data such as county population size, employment status, population below poverty, ways of communicating to work, and health insurance coverage available through the American Community Survey https://www.census.gov/programs-surveys/acs/data.html. In lieu of data regarding travel to ZIKV affected regions (which are not available), we also considered the size of tourism industry for each county. Our full model included 76 factors across four categories (Table 4).

1.1.3 Maximum Entropy Model

Let $X = \{x_1, x_2, ..., x_{254}\}$ represent the 254 counties of Texas. Suppose a case of ZIKV is introduced into Texas and let p_i be the probability that it occurs in county x_i. The vector of p_i sums to one over all counties. We wish to estimate this unknown probability distribution using the historical import data. The relative probabilities $p_1, p_2, ...p_n$ can be constrained with known mean, variance, or other moments of some known $f_j(X)$ for each county. The functions $f_j(X)$ are functions of socioeconomic, environmental, and travel variables in our case (Table 4). Mathematically, we want to:

$$\max_{p_i} \quad \Big(-\sum_{i=1}^{254} p_i \log p_i \Big) \tag{1a}$$

$$\text{s.t.} \quad \sum_{i=1}^{254} p_i f_j(x_i) = E(f_j(X)) \quad \forall j \tag{1b}$$

$$\sum_{i=1}^{254} p_i = 1 \tag{1c}$$

$$p_i \geq 0 \qquad \forall i \tag{1d}$$

When we use Shannon's measure of entropy as the objective (1a), the constraints (1d) are automatically satisfied. The right-hand-side of (1b), $E(f_j(X))$, is estimated by the weighted arithmetic mean of $f_j(x_1), f_j(x_2), ..., f_j(x_{254})$ based on the 254 counties of Texas [4].

1.1.4 Representative Variable Selection

After solving the maximum entropy model, we removed duplicate variables—variables that bring the essentially identical information to the model. Selecting the representative variables was done with a variation of the facility location problem [5]. The goal was to select k variables that adequately represent the entire variable set. The k selected factors would each represent themselves and the remaining $72-k$ variables would be represented by exactly one variable from the k selected variables (Eq 2b). The $\ell-\infty$ norm of the difference between two unit-norm variables, denoted by f_i, f_j in Table 1, was used to measure the distance between pairs of variables. This distance measure was derived from the maximum difference in expectations that the two variables can produce, under any probability distribution. The facility location model allowed us to select the k most representative variables (Eq (2c)). The objective function (Eq 2a) for selecting representative variables was to minimize the distance between the k representative variables and all the variables in the entire variable set.

$$\min_{x_{ij},y_j} \quad \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}x_{ij} \tag{2a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = 1 \qquad \forall i \tag{2b}$$

$$\sum_{j=1}^{n} y_j = k \tag{2c}$$

$$x_{ij} \le y_j \quad \forall i,j \tag{2d}$$

$$x_{ij} \in \{0,1\} \quad \forall i,j \tag{2e}$$

$$y_j \in \{0,1\} \quad \forall j \tag{2f}$$

Symbol	Definition
f_j	72 variables represented by vectors $f_j, j = 1, 2, ...72$
d_{ij}	distance between two variables, measured as $d_{ij} = \|\frac{f_i}{\|f_i\|_2} - \frac{f_j}{\|f_j\|_2}\|_\infty$
x_{ij}	$x_{ij} = 1$ if vector i is represented by vector j; $x_{ij} = 0$, otherwise;
y_j	$y_j = 1$, if vector j is selected as representative vector; $y_j = 0$, otherwise;

Table 1: **Representative variable selection. We first applied this method to reduce our county-level ZIKV importation model from 72 to 20 predictors, and then applied predictive variable selection to reduce it further to 10 predictors.**

1.1.5 Predictive Variable Selection

To further streamline the importation model, we considered several different methods for identifying the most predictive among the selected variables, including hypothesis testing to choose between

nested models [6]. Ultimately, we applied a backward selection approach, outlined in Table 2. In each iteration, the variable that contributed the least to model performance was dropped, until all the variables were eliminated. Specifically, we evaluated model performance through cross validation on out-of-sample data. For each iteration, 12 years DENV importation cases were divided into two subsets: a 9 year training data set (for fitting the model) and a 3 year testing data set (for gauging model performance). We ran each set of variables on 6 randomly selected partitions of the 12 years of available data. From the 6 runs, we calculated the average of the out-of-sample log-likelihood of the model and eliminated the variable that gave the largest mean out-of-sample log-likelihood. Drops in model performance was negligible until fewer than 10 variables were included.

Algorithm	Backward Selection
1	**function** BACKWARD SELECTION (N)
2	Set $V = N$
3	**While** $\|V\| > 1$ **do**
4	Set $e = argmax_{e \in V} C(S(V - e))$
5	Set $V = V - \{e\}$
6	Record V and $C(S(V - e))$
N	The complete set of representative variables
C	Return the out-of-sample log-likelihood, averaged over of seven randomly sampled cross validation folds
S	Fit a maximum entropy model given a set of variables f_j

Table 2: **Predictive variable selection algorithm used to select the 10 most informative variables from among the 20 representative variables selected for the Texas county ZIKV importation model.**

Variables ordered by importance
Total Direct Spending(dollars)
Graduate or professional degree in Percentage
Local (dollars)
Male Population
Commuting to Work with Other Means
Max Temperature of Warmest Month
Population below Poverty Level in Percentage
Precipitation of Wettest Quarter
Population without Health Insurance
Graduate or professional degree population

Table 3: **Import risk model variables**. These 10 variables were selected from 72 variables using a combination of representative variables selection and predictive backwards selection. The importance of each variable (from top to bottom) is determined by order of exclusion in backwards selection, with the most important variables remaining in the model the longest.

Environmental	Socio-economic	Demographic, Travel and Vector Suitability
Annual Mean Temperature	Employed Population	Male Population
Annual Precipitation	Unemployed Population	Female Population
Slope	Employed Population in Percentage	Male Population in Percentage
Population Count	Unemployed Population in Percentage	Female Population in Percentage
Isothermality	Population below Poverty Level in Percentage	Local(dollars)
Precipitation of Driest Month	Families below Poverty Level in Percentage	State(dollars)
Elevation	Population with Health Insurance	Total Direct Spending(dollars)
Maximum Green Vegetation Cover	Percentage with Health Insurance	Visitor Spending
Temperature Seasonality	Population without Health Insurance	Earnings(dollars)
Precipitation Seasonality	Percentage without Health Insurance	Travel Employment
Min Temperature of Coldest Month	Population Walk to Work in Percentage	Average MGV (percentage per km)
Precipitation of Driest Quarter	Percentage Commuting to Work with Taxi	Total Approximate MGV Cover (km)
Max Temperature of Warmest Month	Mean Travel Time to Work(Minutes)	
Precipitation of Wettest Quarter	Population Walk to Work	
Temperature Annual Range	Commuting to Work with Taxi	
Precipitation of Warmest Quarter	Percentage Commuting to Work with Public Transportation	
Mean Temperature of Wettest Quarter	Commuting to Work with Public Transportation	
Precipitation of Coldest Quarter	Commuting to Work with Car, Truck or Van (Carpooled)	
Mean Temperature of Driest Quarter	Commuting to Work with Car, Truck or Van(Alone)	
Mean Temperature of Warmest Quarter	Percentage Commuting to Work with Car, Truck or Van(Carpooled)	
Mean Temperature of Coldest Quarter	Percentage Commuting to Work with Car, Truck or Van(Alone)	
Mean Diurnal Range	Commuting to Work with Other Means	
Precipitation of Wettest Month	Percentage Commuting to Work with Other Means	
Aspect	Education Attainment below 9th grade	
Artificial Surface Cover(Percentage)	Education Attainment below 9th grade in Percentage	
Total Artificial Surface Cover (km)	Education Attainment between 9th and 12th grade	
	Percentage Education Attainment between 9th and 12th grade	
	High School Graduates	
	High School Graduates in Percentage	
	College without diploma	
	College without diploma in Percentage	
	Associates degree	
	Associates degree in Percentage	
	Bachelor's degree	
	Bachelor's degree in Percentage	
	Graduate or professional degree	
	Graduate or professional degree in Percentage	

Table 4: **All 72 environmental, socioeconomic, and behavioral variables evaluated for Texas county ZIKV importation model**

1.2 County R_0 Estimation

Following a ZIKV introduction, epidemic emergence requires sufficient vector-human interactions to maintain an autochthonous transmission cycle. To understand the potential for this second component of emergence, we estimated the reproduction numbers (R_0) for Texas counties following the methodology in [7] and according to the Ross-Macdonald formulation, given by,

$$R_0 = \frac{mbc\alpha^2 e^{-\mu\omega}}{\mu\gamma}, \tag{3}$$

with the variables defined in Table 5.

Parameter	Description	Value	Reference
m	Mosquito abundance. Variation in intrinsic suitability and GDP across counties	0.006-0.844	[8, 7]
b	Mosquito-to-human probability of transmission per bite	0.634 [0.214, 0.8]	[9]
c	Human-to-mosquito probability of transmission per bite	0.77 [0.6, 0.95]	[10]
α	Mosquito biting rate: the expected number of bites per day.	0.63 [0.4, 0.8]	[11]
μ	Average lifespan of female *Ae. aegypti* mosquito (days)	14	[12]
ω	Extrinsic incubation: The expected days between initial infection and infectiousness in *Ae. aegypti*. Variable according to average county temperature.	6-18	[13]
γ	Rate of recovery out of infectious period (1/days)	1/9	[14]

Table 5: **Parameters for estimating ZIKV R_0 in Texas counties. Bounds in uncertainty in estimated parameters from original sources are included. The ranges in extrinsic incubation and mosquito abundance are Texas specific and reflect those used in the analysis.**

.

As both the mosquito lifespan (μ) and the extrinsic incubation period (ω) for ZIKV change with temperature, we estimated average August temperatures from 1980 to 2010 for each county [15], which ranged from 24 to 31 °C. To estimate temperature-dependent extrinsic incubation periods, we used the log-normal distribution model estimated in [13] for DENV viruses in *Ae. aegypti*. Although μ does vary with temperature, a distribution based on field mark-release-recapture studies of *Ae. aegypti* estimated that adult longevity stays roughly the same over the range of average August temperatures across Texas. Therefore we assumed a single value across Texas, using the average value of adult longevity where 50% of the population remained. This decision is also consistent

with findings that there is less variation in *Ae. aegypti* survival at the mid-range temperatures seen in Texas [12].

In addition to temperature dependent effects on ω, the mosquito abundance (m) varies among Texas counties. We used estimates of occurrence probabilities of *Ae. aegypti* for each Texas county obtained from a predicted global distribution of *Ae. aegypti* in [8]. We summed the $5km^2$ occurrence probabilities for all cells within a county to get the occurrence probability of *Ae. aegypti* for a county. Following the methodology in [7], we assumed mosquito abundance is distributed as a Poisson variable, and therefore the probability that there is at least one mosquito in a county follows the relationship $1 - exp(-\lambda)$ [16]. Using the inverted equation $\lambda = -ln(1 - occurrence\ probabillity)$, where λ is the expected abundance of mosquitoes, we obtained a proxy for the mosquito abundance of each county. We further multiplied the mosquito abundance by a log linear function of the 2014 gross domestic product economic index [17] for each Texas county extended from the fitted function derived in [7], to incorporate economic effects on mosquito-human contacts (see fig 6, *Expected*). This was done to account for characteristics of higher economic locations that would reduce contact with mosquitoes. In this way, m variation among counties comes from both global suitability estimates and GDP effect.

We used recent data on the viral infection rate of Brazilian populations of *Ae. aegypti* to the currently circulating Asian genotype of ZIKV to estimate the human-to-mosquito transmission probability [10]. We derived an estimate of the mosquito-to-human transmission probability parameter using the median estimate of the published fitted parameter for the transmission rate, $\beta_h = 0.4$, from [9], and an estimate of the *Ae. aegypti* biting rate, $\alpha = 0.63$ from [11]. The transmission rate term, β_h, is a product of the biting rate and the mosquito-to-human transmission probability, b. Therefore we derived the mosquito-to-human transmission probability parameter, using the equation $b = \beta_h/\alpha$, giving $b = 0.634$ (5) .

We explore the uncertainty in our county R_0 estimates further below.

2 ZIKV Outbreak Simulations

2.1 Simulating Outbreaks

We model ZIKV outbreaks using a stochastic Markov branching process model (Fig 2). To transmit ZIKV, a mosquito must bite an infected human, get infected with the virus, bite a susceptible human, and the human must then get infected with the virus. Rather than explicitly model the full transmission cycle, we aggregate the two-part cycle of ZIKV transmission (mosquito-to-human and human-to-mosquito) into a single exposed class, and do not explicitly model mosquitoes. For the purposes of this study, we need only ensure that the model produces a realistic human-to-human generation time of ZIKV transmission.

Our simulations begin with a single ZIKV imported case, and we simulate the Susceptible-Exposed-Infectious-Recovered (SEIR) transmission process that follows. The temporal evolution of the compartments are governed by daily probabilities for infected individuals transitioning between E, I and R states, new ZIKV introductions and transmissions, and reporting of current infectious cases (Table 6). We assume that infectious cases cause a Poisson distributed number of secondary cases per day (via human to mosquito to human transmission), and that low reporting rates correspond to the percentage ($\sim 20\%$) of symptomatic ZIKV infections [18]. Although reporting has been as low as 10% in historical ZIKV outbreaks, we focus primarily on 20% reporting for the majority of results. We make the simplifying assumption that asymptomatic cases transmit ZIKV at the same rate as symptomatic cases, which can be modified if future evidence suggests otherwise.

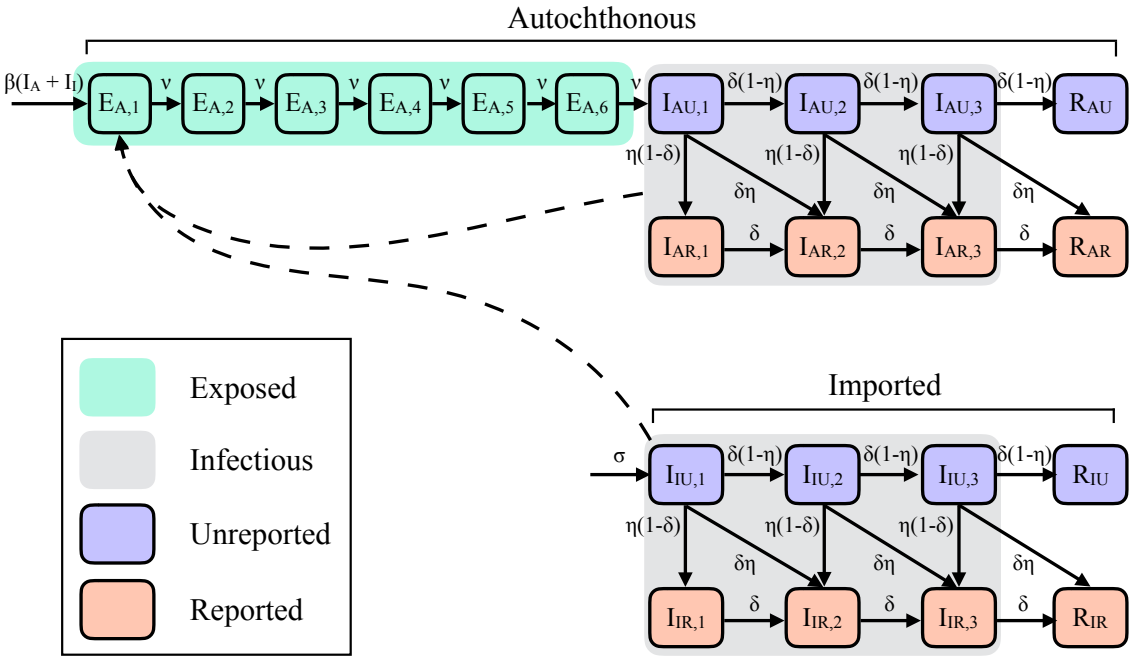

Figure 2: **Diagram of ZIKV outbreak model.** The model tracks disease progression, transmission, and reporting of both imported and autochthonous ZIKV cases. Individuals progress through compartments via a daily Markovian process, according to the solid arrows in the diagram. The *Exposed* and *Infectious* periods consist of several (boxcar) compartments to simulate realistic outbreak timing. Unreported infected individuals have a daily probability of being reported. Imported cases are assumed to arrive daily according to a Poisson distribution (with mean σ) at the beginning of their infectious period, and otherwise follow the same infectious process as autochthonous cases. Autochthonous transmission occurs at rate $\beta(I_A + I_I)$, where I_A and I_I are the total number of infectious autochthonous and imported cases, respectively (dashed arrows).

We give the model equations below for both introduced cases (Eqs 4) and autochthonous cases (Eqs 5).

For each scenario, consisting of a particular importation rate, transmission rate, and reporting rate, we ran 10,000 stochastic simulations. Each simulation began with a single infectious unreported importation and terminated when there were no individuals in either the *Exposed* or *Infectious* classes or the cumulative number of autochthonous infections reached 2,000. We classified simulations as either epidemics or self-limiting outbreaks; epidemics are those reaching 2,000 cumulative autochthonous infections with a maximum daily prevalence of at least 50 (Fig 3). We define daily prevalence as the number of current unreported and reported autochthonous infections.

2.2 Model Equations: Introduced ZIKV Cases

Unreported compartments:

$$I_{IU,1}(t) = I_{IU,1}(t-1) + Pois\,(x = 1, p = \sigma) - (\delta + \eta - \delta\eta) I_{IU,1}(t-1) \tag{4a}$$

For $i \in 2, \ldots, n$ infectious compartments:

$$I_{IU,i}(t) = I_{IU,i}(t-1) + \delta(1-\eta)I_{IU,i-1}(t-1) - (\delta + \eta - \delta\eta)I_{IU,i}(t-1) \tag{4b}$$
$$R_{IU}(t) = R_{IU}(t-1) + \delta(1-\eta)I_{IU,n}(t-1) \tag{4c}$$

Reported compartments:

$$I_{IR,1}(t) = I_{IR,1}(t-1) + \eta(1-\delta)I_{IU,1}(t-1) - (\delta)I_{IR,1}(t-1) \tag{4d}$$

For $i \in 2, \ldots, n$ infectious compartments:

$$\begin{aligned} I_{IR,i}(t) = I_{IR,i}(t-1) &+ (\delta)I_{IR,i-1}(t-1) + (\delta\eta)I_{IU,i-1}(t-1) + \eta(1-\delta)I_{IU,i}(t-1) \\ &- (\delta)I_{IR,i}(t-1) \end{aligned} \tag{4e}$$
$$R_{IR}(t) = R_{IR}(t-1) + (\delta)I_{IR,n}(t-1) + (\delta\eta)I_{IU,n}(t-1) \tag{4f}$$

2.3 Model Equations: Autochthonous ZIKV Cases

Exposed compartments:

$$\begin{aligned} E_{A,1}(t) = E_{A,1}(t-1) &+ \sum \left(Pois\left(x = \sum_{j \in \{IU,IR,AU,AR\}} \sum_{k=1}^{n} I_{j,k},\ \ p = \beta \right) \right) \\ &- (\nu)E_{A,1}(t-1) \end{aligned} \tag{5a}$$

For $i \in 2, \ldots, e$ exposed compartments:

$$E_{A,i}(t) = E_{A,i}(t-1) + (\nu)E_{A,i-1}(t-1) - (\nu)E_{A,i}(t-1) \tag{5b}$$

Unreported compartments:

$$I_{AU,1}(t) = I_{AU,1}(t-1) + (\nu)E_{A,e}(t-1) - (\delta + \eta - \delta\eta)I_{AU,1}(t-1) \tag{5c}$$

For $i \in 2, \ldots, n$ infectious compartments:

$$I_{AU,i}(t) = I_{AU,i}(t-1) + \delta(1-\eta)I_{AU,i-1}(t-1) - (\delta + \eta - \delta\eta)I_{AU,i}(t-1) \tag{5d}$$
$$R_{AU}(t) = R_{AU}(t-1) + \delta(1-\eta)I_{AU,n}(t-1) \tag{5e}$$

Reported compartments:

$$I_{AR,1}(t) = I_{AR,1}(t-1) + \eta(1-\delta)I_{AU,1}(t-1) - (\delta)I_{AR,1}(t-1) \tag{5f}$$

For $i \in 2, \ldots, n$ infectious compartments:

$$\begin{aligned} I_{AR,i}(t) = I_{AR,i}(t-1) &+ (\delta)I_{AR,i-1}(t-1) + (\delta\eta)I_{AU,i-1}(t-1) + \eta(1-\delta)I_{AU,i}(t-1) \\ &- (\delta)I_{AR,i}(t-1) \end{aligned} \tag{5g}$$
$$R_{AR}(t) = R_{AR}(t-1) + (\delta)I_{AR,n}(t-1) + (\delta\eta)I_{AU,n}(t-1) \tag{5h}$$

With $Pois\,(x, p)$ indicating x random draws from a Poisson distribution with $\lambda = p$, subscripts of I and A, respectively, indicating introduced and autochthonous cases, subscripts of R and U, respectively, indicating reported and unreported cases, and parameter values defined as described in Table 6.

Parameter	Description	Values Investigated (or median 95%)	Source
Exposed compartments (e)	Number of exposed compartments	6	[14, 19]
Incubation Rate (ν)	Daily probability of progressing from one exposed compartment to the next	0.584	[14, 19]
Infectious compartments (n)	Number of infectious compartments	3	[14, 19]
Recovery Rate (δ)	Daily probability of progressing from one infectious compartment to the next	0.3041	[14, 19]
Reproduction Number (R_0)	The expected total number of secondary infections from one infectious individual in a fully susceptible population	0-2.2	County R_0 estimates
Daily Reporting Rate (η)	The daily probability of an infectious individual being reported	Daily: $0.011 - 0.0224$ Overall: $10 - 20\%$	[18]
Daily Importation Rate (σ)	The expected number of infectious ZIKV importations per day	$0.0 - 1.21$	County importation rate estimates
Generation Time	The average length of time between consecutive exposures $GT = \frac{e}{\nu} + (\frac{1}{2})\frac{n}{\delta}$	15 (9.5-23.5) days	[19]

Table 6: **Stochastic ZIKV outbreak model parameters**. We hold the disease progression parameters constant across all scenarios, estimate R_0 and importation rate for each individual county, and vary the reporting rate to investigate its impact on the uncertainty of ZIKV risk assessments.

2.4 Fitting the Generation Time

To capture the correct outbreak timing, we fit the generation time of our SEIR model to estimates for the ZIKV exposure and infectious periods in humans. The generation time measures the average duration from initial symptom onset to the subsequent exposure of a secondary case, and is estimated to range from 10 to 23 days for ZIKV [16]. In our model, the generation time corresponds to the sum of the exposure period and 1/2 the infectious period. We therefore fit the infectious period in our model to human ZIKV estimates for duration of viral shedding, and then fit the exposure period so that the sum of the two classes match the estimated ZIKV serial interval.

According to our modeling framework: with one infectious compartment, the distribution of waiting times in the compartment would follow a geometric distribution, with the most common waiting time equal to one day regardless of the transition rate. As this is a biologically unrealistic waiting time distribution, we use Boxcar implementations to yield a more realistic distribution [20]. In such a framework one splits a compartment into multiple separate compartments (boxes), has individuals transition through these compartments, and alters the transition rate for each compartment so the average waiting time spent in all compartments equals that of the original desired average. For example, if a 10 day infectious period were desired, one could model the infectious period as 1 compartment with a daily transition rate of 1/10, or 5 compartments with a daily transition rate of 5/10. The number of infectious individuals is either the number of individuals in the single compartment, or the total number of individuals in all five boxes. Both scenarios would have an average waiting time of 10 days to move through the infectious period, but the 5 boxes would necessitate individuals being infectious for at least 5 days giving a more realistic waiting time distribution that follows a negative binomial distribution (sum of multiple independent geometric distributions).

First, we solved for transition rates and compartments of a Boxcar Model infectious period that yielded an infectious period with 3 compartments and mean duration of 9.88 days and 95% CI of (3-22) [14]. Then, we fit the exposure period so that the combined duration of the infectious and exposure periods matched the empirical ZIKV generation time range [19], yielding 6 compartments and a mean exposure period of 10.4 days (95% CI 6-17) and finally a mean generation time of 15.3 days (95% CI 9.5-23.5). Given that the exposure period includes human and mosquito incubation periods and mosquito biting rates, this range is consistent with the estimated 5.9 day human ZIKV incubation period [14]

3 Risk Assessment and Surveillance Trigger Analysis

Although ZIKV surveillance data will ultimately be used for many planning and response purposes, here we focus on just one: assessing the potential for epidemic expansion. This is intended as a demonstration and test of the approach, which can be similarly applied to plan and improve other surveillance activities.

We classify simulations as epidemics if the reaches at least 2,000 autochthonous cases and, at least once, surpass a daily autochthonous prevalence of 50. The second criteria was systematically designed to distinguish (1) simulations with high transmission rates from (2) simulations with low transmission rates but high importation rates (Fig 3). The daily prevalence threshold of 50 ensures that the vast majority of outbreaks and epidemics are classified correctly. Occasionally, outbreaks with $R_0 \leq 1$ and a high importation rate grow sufficiently large to be classified as epidemics, even though they technically are not. Since they would warrant substantial a public health response [21], we let the classification stand. As discussed in the main text, such misclassifications arise only under exceedingly high importations rates and do not qualitatively influence our results.

To find the epidemic risk in a county upon seeing, x, reported cases, we first find all trials in our 10,000 simulations that encounter x reported cases, and then find what proportion of those simulations become an epidemic. For example, if 1,000 of a county's simulated outbreaks have 2 reported cases, but only 50 of those simulations become epidemics, than the epidemic risk upon seeing 2 cases in that county would be 5%. This framework allows us to assess 1) the likelihood of a county experiencing x reported cases and 2) the probability of sustained transmission upon a second reported cases (assuming no subsequent intervention).

We only analyze triggers for counties where at least 1% of simulations reach the trigger value (number of reported cases), to avoid accidentally inflating the risk of counties that have only very few simulations reaching the trigger value. This method also does not distinguish between counties with 500 epidemics out of 1,000 *triggered* simulations from those with 5,000 of 10,000. Thus, we report the probability of a *triggered* outbreak separately. Consider a county with an $R_0 = 1.1$ and another county with a much higher R_0. In the second county, outbreaks are much more likely to progress into epidemics. However, both counties should interpret a cluster of reported cases as strong indication of epidemic expansion, regardless of the prior probability that such a cluster would occur.

We evaluate epidemic risk across Texas counties following two reported autochthonous cases, in line with recent CDC's guidelines [21]. As demonstrated in the main text, our framework can also be applied to design surveillance triggers, based on local epidemic risks and reporting rates. We show our full Texas risk assessment under a worse case elevated importation scenario in Fig 5.

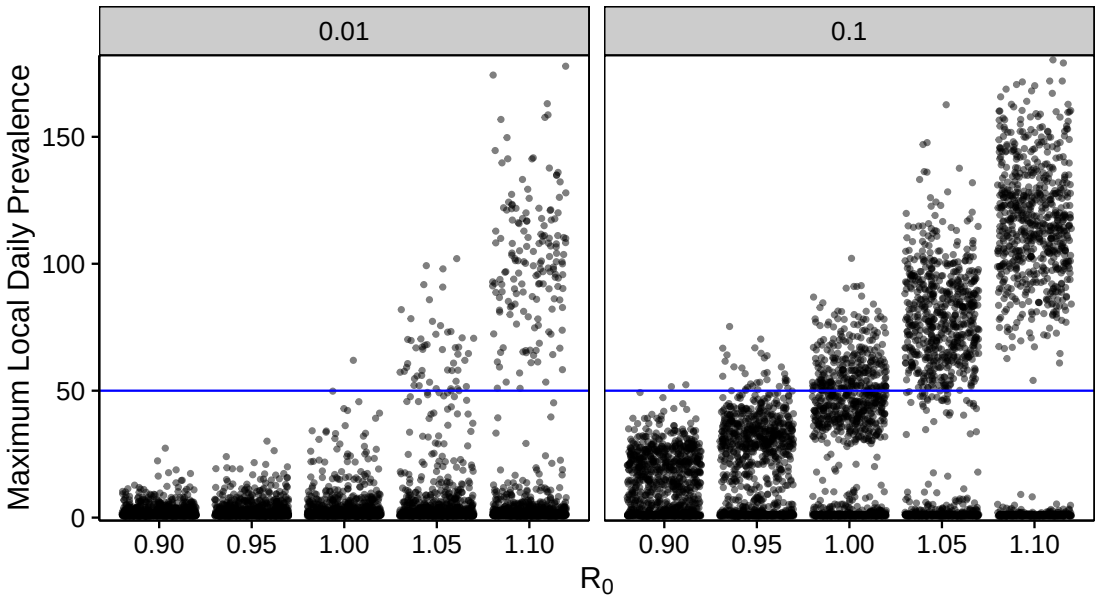

Figure 3: **Selecting daily prevalence threshold for distinguishing self-limiting outbreaks from epidemics.** Across a range of R_0 values, we plot the maximum daily total autochthonous infectious individuals for 1,000 of our 10,000 trials (black dots). The blue line indicates the threshold (50) selected to differentiate epidemics with $R_0 > 1$ from outbreaks with $R_0 \leq 1$. At a low importation rate (0.01), the majority of simulations with $R_0 \leq 1$ are self-limiting and rarely progress into large sustained outbreaks. As R_0 increases, a greater proportion of simulations exceed the threshold. As the importation rate increases (panels from left to right) the separation between self-limiting outbreaks and epidemics becomes more pronounced.

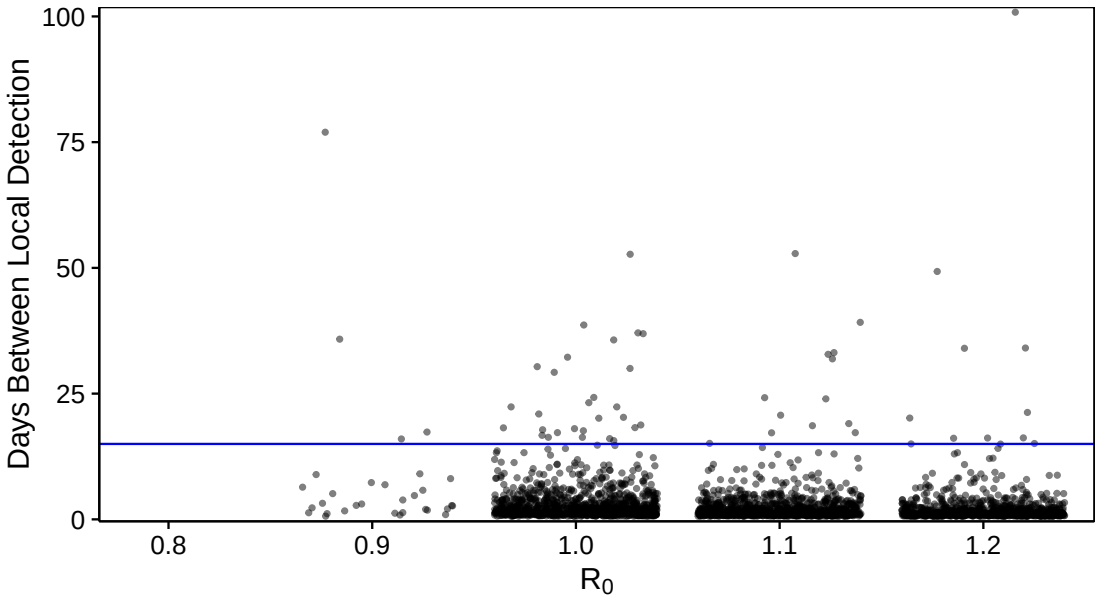

Figure 4: **Time between detection of locally transmitted cases during epidemics.** Across a range of R_0 values with an importation rate 0.1 cases/day, we plot the time between detection events of autochthonous cases for simulations out of the 10,000 trials in which epidemics occurred (black dots). The blue line indicates a two-week threshold as recommended by the CDC for follow-up of local transmission. Even under a high importation rate of 0.1 cases/day, epidemics do not occur when $R_0 = 0.8$, and rarely occur when $R_0 = 0.9$. As R_0 increases, a greater proportion of simulations have fewer days in between detection events as the number of infections rapidly increase.

3.1 Texas risk assessment - Worst case importations

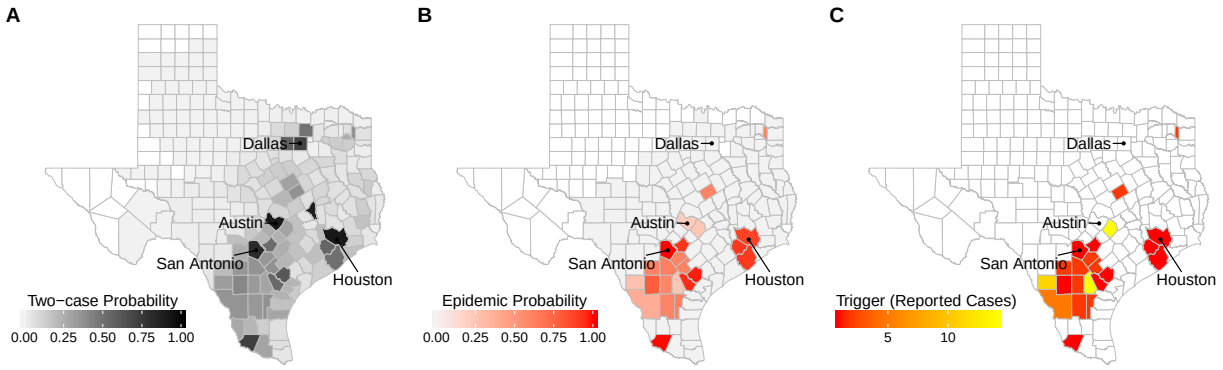

Figure 5: **Texas risk assessment under high importation rate.** (A) Probability of detecting two ZIKV cases in a county assuming a 20% reporting rate and an importation scenario for August 2016 that assumes that recently observed ZIKV importations account for only 20% of actual importations (405 cases statewide per 90 days). (B) Under the same scenario as (A), the probability of an impending epidemic at the moment the second ZIKV case is reported. White counties never reach two reported cases, across all 10,000 simulated outbreaks; light gray counties reach two cases, but never experience epidemics. (C) Recommended county-level surveillance triggers for detecting that the probability of an epidemic has exceeded 50%, assuming a reporting rate of 20% and the elevated importation scenario. White counties indicate that fewer than 1% of the 10,000 simulated outbreaks reached two reported cases.

4 R_0 Uncertainty Analysis

We assessed the sensitivity of our county-level R_0 estimates to several key epidemiological assumptions about which there is significant uncertainty. In all cases, we find that our original estimates robustly capture relative ZIKV transmission risks among counties, but not necessarily the absolute risks. We separate the sensitivity analysis into factors that vary across counties and universal parameters that do not.

4.1 Sensitivity Analysis: County-specific inputs

We separately consider two primary sources of variation in the R_0 estimates across Texas counties: (1) mosquito-human contact rates and (2) temperature-dependent extrinsic incubation periods (EIPs)

Varying the impact of GDP on human-mosquito contact rates We estimated *Ae. aegypti* abundance using the predicted global distribution published in [8], and the method proposed in [7] assuming the mosquito abundance, m, follows a Poisson distribution [16]. As was done in [7], we multiplied m for each county by a scalar factor (less than one) that reduces human-mosquito encounters as a function of socioeconomic status, as estimated by the 2014 gross domestic product economic index (GDP) [17]. Specifically, we apply the log-linear function proposed in [7] to log-GDP's in Texas, which range from $\sim$ 10 to 12 on a natural log scale ($\sim$ \$23,000 to $\sim$ \$97,000).

Consequently, counties with higher GDP have lower estimated R_0's. However, the shape of the GDP to m function has uncertainty associated with it, and has not been estimated for Texas specifically. Here, we consider four different relationships between GDP and the R_0 scaling factor, m, to account for uncertainty in estimating county local transmission risk (Fig 6).

- The *Expected* or baseline scenario, is based on the average of the fitted function in [7]. The model assumes a negative log-linear relationship between GDP and scalar multiple of human-mosquito interactions.
- The *Stronger* scenario assumes that, as a starting condition, R_0 is even more sensitive to GDP relative to baseline, but decreases at the same rate for every unit increase in GDP as in the baseline. Under this scenario, human-mosquito interactions (and thus R_0) tend to be diminished throughout Texas, given Texas' higher socioeconomic conditions relative to ZIKV affected regions throughout the Americas.
- The *Weaker* scenario assumes that GDP does not strongly impact human-mosquito interaction, and mosquito to human ratios in each county closely reflect the mosquito abundances estimated from environmental predictions. However, the scalar multiple decreases at the same rate for every unit increase in GDP as in the *Expected* and *Stronger* scenarios. Under this scenario, human-mosquito interactions (and thus R_0) in Texas are assumed to be similar to those found in other ZIKV affected regions, despite the better socioeconomic conditions in Texas.
- The *Heterogeneous* scenario assumes a greater reduction in human-mosquito interactions per unit gain in GDP compared to the first three scenarios and assumes a starting condition similar to the *Expected* scenario. Counties with low GDP experience levels of human-mosquito interactions similar to those found in other ZIKV affected regions and have high R_0s. However, counties with higher GDPs have a greater reduction in human-mosquito interactions than in *Stronger* and consequently lower R_0s.

The first three scenarios simply scale the magnitude of R_0s, but do not change the relative transmission risks between the counties (Fig 7). Based on the limited known historical arbovirus activity in Texas, the *Expected*, *Stronger*, and *Heterogeneous* scenarios may be the most likely. Other than limited outbreaks of DENV along the Texas-Mexico border in areas predicted to have $R_0 s > 1$, Texas has experienced minimal DENV or CHIKV transmission. This suggests the *Weaker* scenario assumed relationship is unlikely, and supports the idea that socioeconomics substantially influence arbovirus transmission potential. The impact of socioeconomics on human-mosquito interaction may be critical to anticipating future arbovirus emergence, and warrants future investigation.

Varying temperature The extrinsic incubation periods of ZIKV in *Ae. aegypti* are temperature dependent. Here, we consider seasonal changes in this factor and its impact on the estimated R_0. Specifically, we calculate monthly R_0 estimates throughout the summer and fall of 2016, based on historical monthly average temperatures (from 1980-2010) for each county [15] (Fig 8). Transmission risk is expected to be stable throughout the summer, and into September. In October, however, the number of counties with $R_0 \geq 1$ is expected to decline to zero. However the relative risk of the counties also changes in October, with a strong movement of risk towards the southern (warmer) counties in the state.

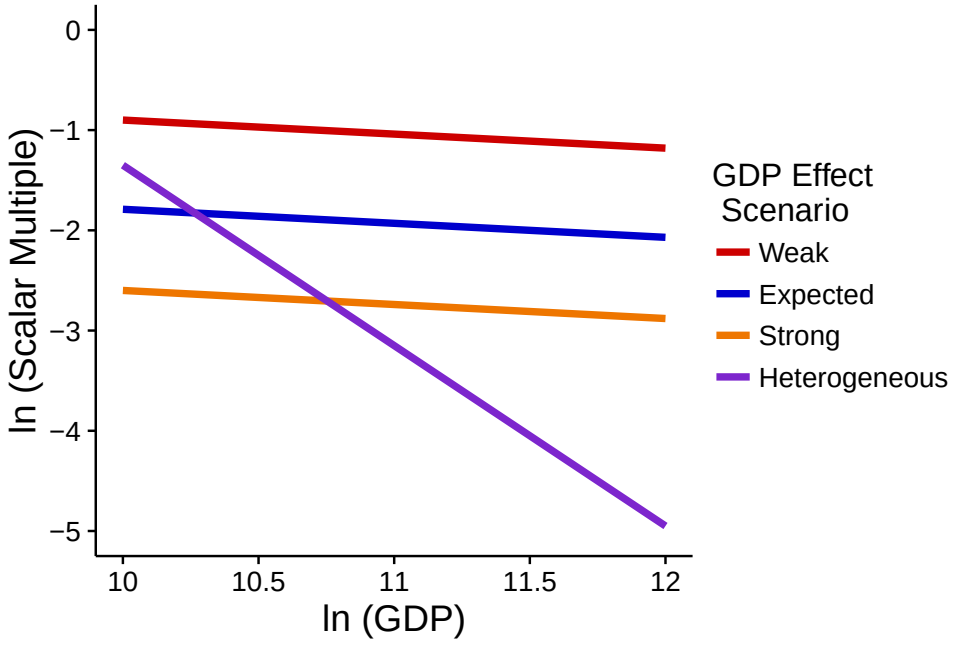

Figure 6: **Four models relating GDP to the human-mosquito interaction rate.** Each model specifies a declining relationship between GDP and the scalar multiple used to reduce human-mosquito contact rate when estimating each county R_0. In all cases, higher GDP translates to lower R_0. Note, both axes are on a log scale.

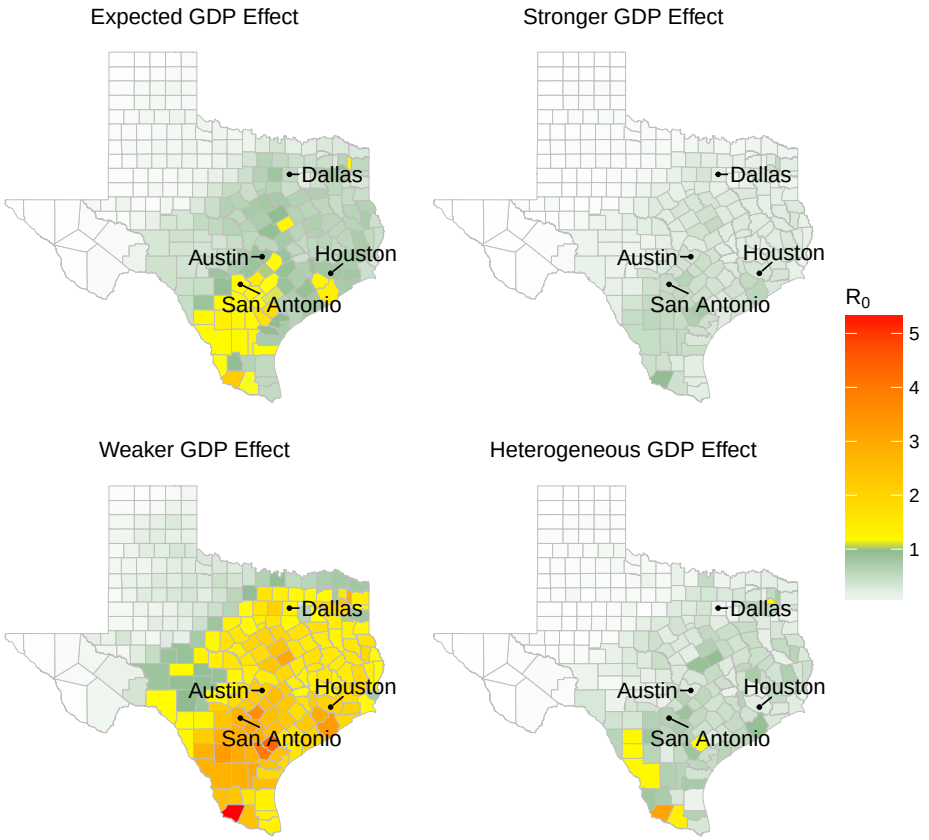

Figure 7: **Estimated R_0 under four different socioeconomic impact scenarios.** The *Expected* scenario corresponds to our original estimates of R_0 given in the main text. The *Stronger* scenario assumes that the impact of GDP on mosquito-human contact is greater (on reducing contact) than under the *Expected* scenario; no counties have an estimated $R_0 > 1$. Under the *Stronger* scenario, statewide GDP is assumed to be sufficiently high to limit ZIKV transmission and only one county remains at moderate risk. The *Weaker* scenario assumes that the effect of GDP is minimal; R_0 estimates are approximately two-fold higher than under the *Expected* scenario, including 46 counties with estimated $R_0 > 2$. The highest risks occur in Eastern and Southern Texas. Under the *Heterogeneous* scenario, the number of counties with estimated $R_0 > 1$ is fewer than under the *Expected* scenario but greater than under the *Stronger* scenario, with only six counties having an $R_0 \geq 1$.

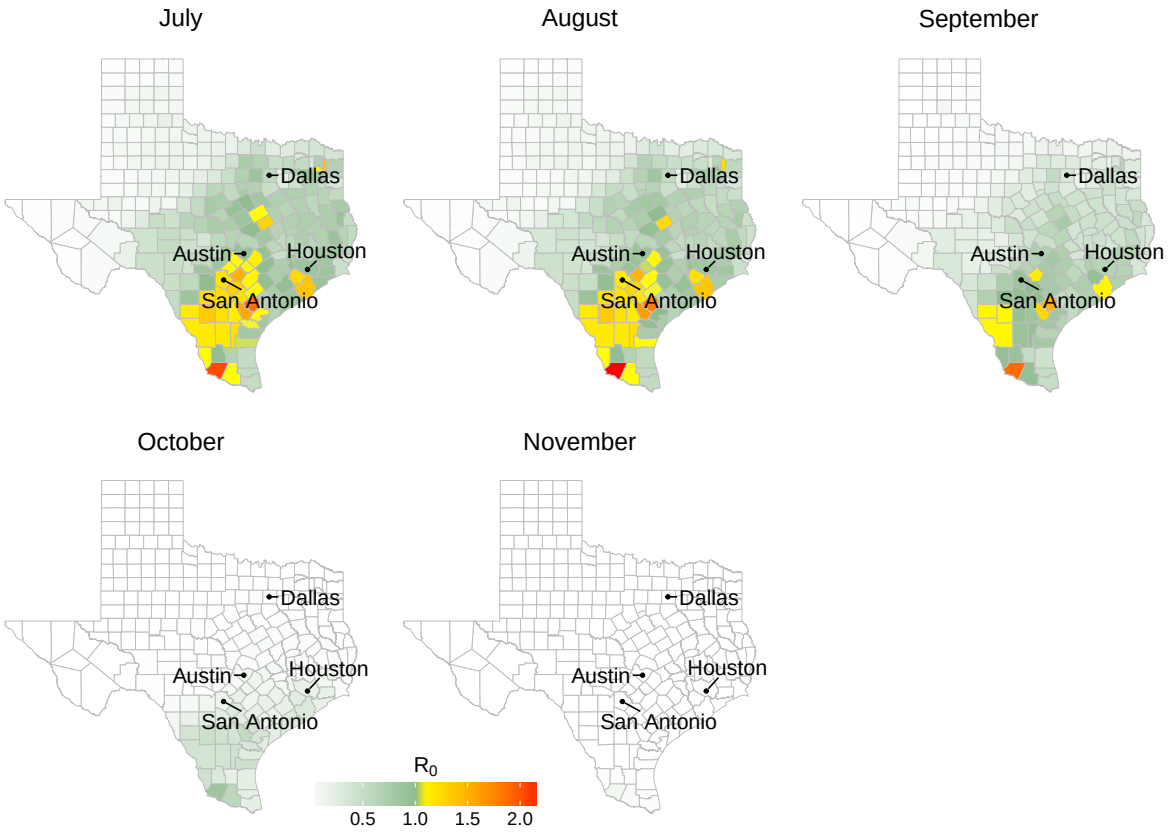

Figure 8: **Monthly R_0 estimates based on changes in the temperature-dependent extrinsic incubation period of ZIKV in *Ae. aegypti*.** This assumes the baseline GDP scenario described above.

4.2 Sensitivity Analysis: Universal Inputs

We measured the sensitivity of our R_0 estimates to variation in parameters that are assumed constant throughout the state, specifically the mosquito biting rate (α), the mosquito-human transmissibility (b), and the human-mosquito transmissibility (c) (Figs 9 - 11). According to Eq. 3, the estimated R_0's will change in the same direction as each of these parameters. Thus, we considered the upper and lower bound estimates of each parameter according to their original sources (Table 5).

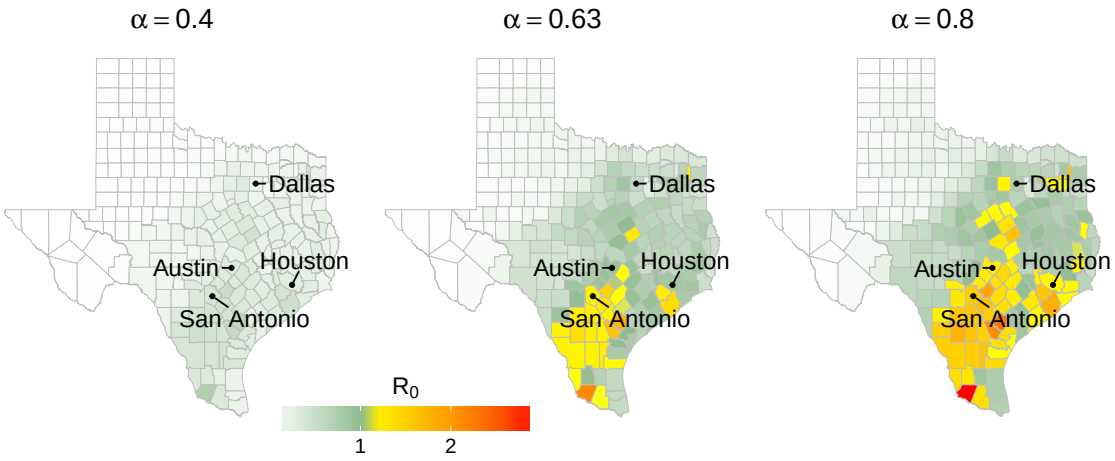

Figure 9: **Sensitivity of estimated R_0 to the biting rates (α).** Our baseline model assumes $\alpha = 0.63$, as estimated from feeding rates of *Ae. aegypti* in Puerto Rico [12]. A $\sim 25\%$ increase in biting rate results in an average increase of 0.15 in estimated R_0, and brings 35 more counties into the high risk category of $R_0 > 1$. Under this scenario, the maximum R_0 estimate is 2.9

4.3 Summary of R_0 Uncertainty Analysis

We compare these analyses using the number of counties that are expected to have $R_0 > 1$ (Fig. 12). Depending on the assumptions, the expected proportion of at risk counties ranges from 0% to 55%. R_0 appears most sensitive to assumptions regarding the relationship between GDP and human-mosquito contact patterns. Temperature can also have a substantial impact, with the maximum risk expected to occur in July and dropping significantly by October. Finally, increasing universal parameters (the mosquito biting rate, and the mosquito-human and human-mosquito transmission probabilities) to their estimated upper bounds increases the expected number of counties at risk by a factor of two to three. Given the lack of significant DENV/CHIKV epidemics in Texas historically, our ZIKV R_0 estimates may be slightly elevated. However, without additional data or analysis, we believe that the baseline county-level R_0 estimates reported in the main text for August and the accompanying temperature-dependent estimates are reasonable, but should only be applied with full consideration of the underlying uncertainty.

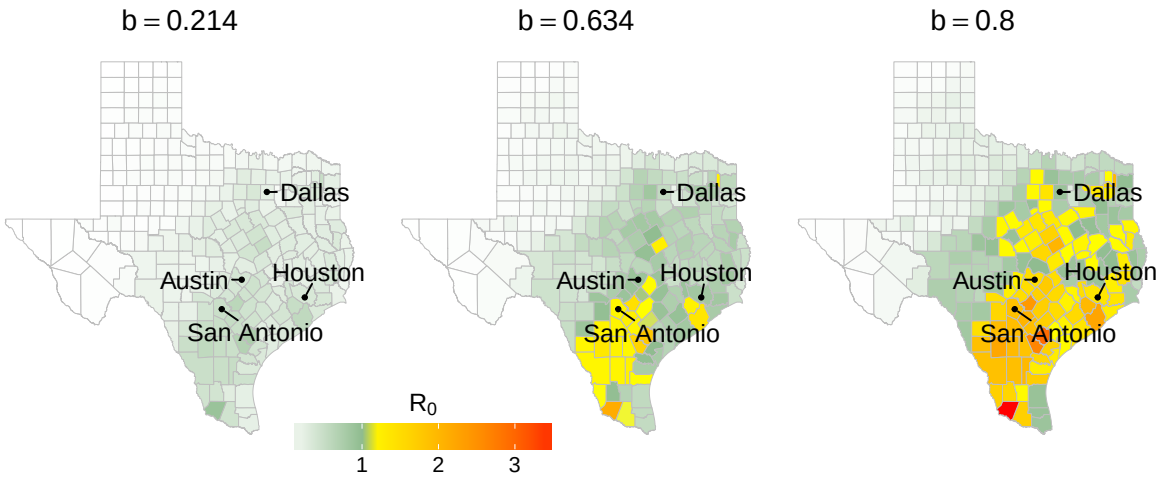

Figure 10: **Sensitivity of estimated R_0 to mosquito-to-human probability of transmission per bite (b).** A $\sim$ 25% increase in b yields an average increase of 0.32 in R_0. The number of counties expected to be at high risk increases three-fold to 91, with the highest R_0 estimated at 3.6. A substantial decrease (35%) in transmission probability would reduce the number of high risk counties to zero.

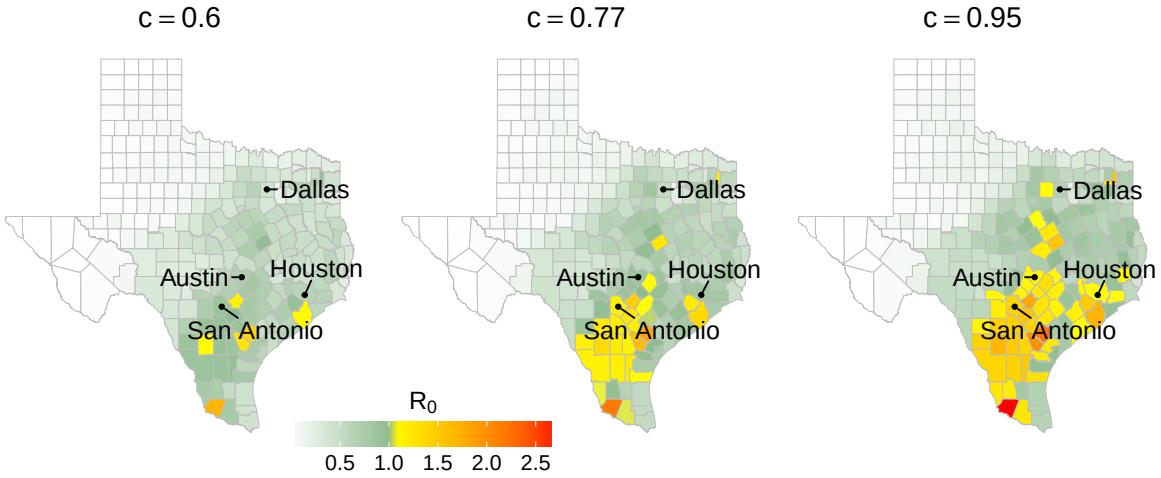

Figure 11: **Sensitivity of estimated R_0 to the human-mosquito transmission probability (c).** A $\sim$ 25% increase in c yields an average increase of 0.11 in R_0. Although this almost doubles the numbers of counties that are considered high risk, this parameter has a lesser impact on R_0 than α and b.

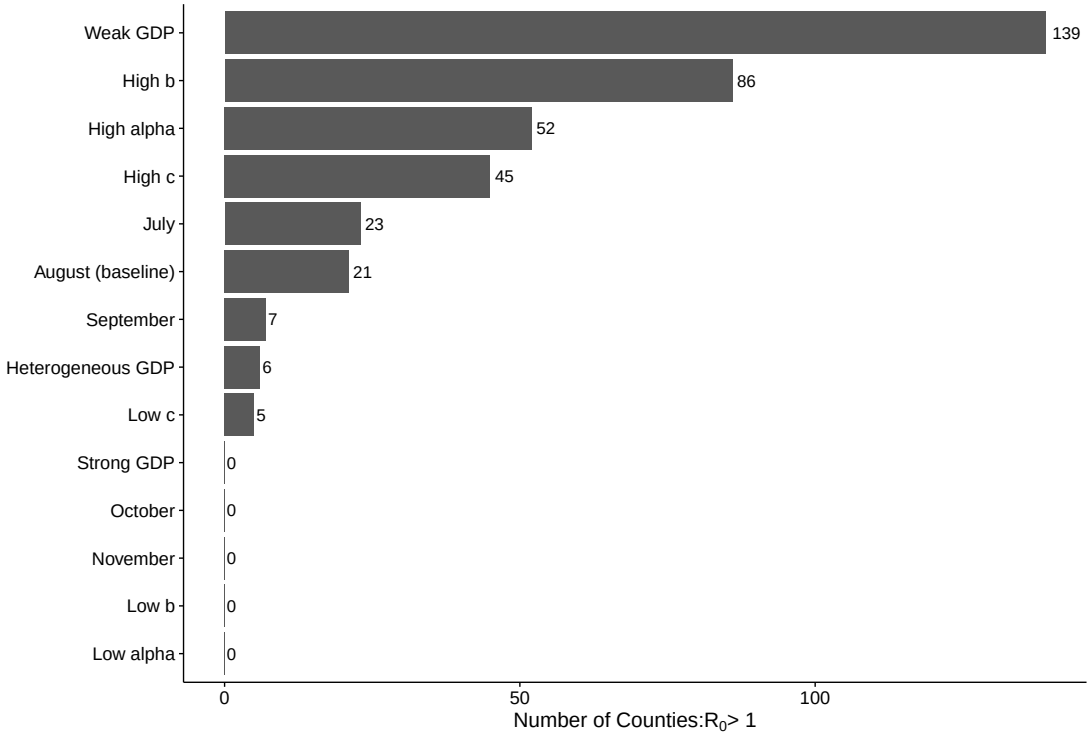

Figure 12: **Comparing sensitivity of R_0 to different county-specific and universal parameters.** The number of counties with R_0 above 1 for each of the 14 sensitivity analysis scenarios described above. The baseline scenario described in the manuscript is for *August*, with 21 counties estimated to have R_0s greater than 1.

References

[1] Texas Department of State Health and Human Services. Arbovirus Activity in Texas 2013 Surveillance Report. Texas Department of State Health Services; Infectious Disease Control Unit Zoonosis Control Branch; 2013. Available from: https://www.google.com/url?sa=t{&}rct=j{&}q={&}esrc=s{&}source=web{&}cd=2{&}ved=0ahUKEwiE8OzU0I{_}OAhVG5SYKHTBTA9QQFggjMAE{&}url=http{%}3A{%}2F{%}2Fwww.dshs.texas.gov{%}2FIDCU{%}2Fdisease{%}2Farboviral{%}2Fwestnile{%}2FReports{%}2F2013WNVannual.doc{&}usg=AFQjCNGhtdb9UNmhhGUpuomsZoqnIng90g{&}sig2=lco86tadgGbKPHa3EK24qg{&}cad=rja.

[2] Texas Department of State Health and Human Services. Arbovirus Activity in Texas 2014 Surveillance Report. Texas Department of State Health Services-Infectious Disease Control Unit Zoonosis Control Branch; 2014. Available from: https://www.google.com/url?sa=t{&}rct=j{&}q={&}esrc=s{&}source=web{&}cd=3{&}ved=0ahUKEwiE8OzU0I{_}OAhVG5SYKHTBTA9QQFggoMAI{&}url=https{%}3A{%}2F{%}2Fwww.dshs.texas.gov{%}2FWorkArea{%}2Flinkit.aspx{%}3FLinkIdentifier{%}3Did{%}26ItemID{%}3D8589999503{&}usg=AFQjCNGO9cW43{_}oOrfmhjWIKseBh9Q1n9g{&}sig2=nfqHpn1nm50ns6fDVBMgMg{&}cad=rja.

[3] Texas Arbovirus Risk. The University of Texas:Scholar Works; 2015. Available from: https://repositories.lib.utexas.edu/handle/2152/31934.

[4] Kapur JN, Kesavan HK. Entropy optimization principles with applications. Academic Pr; 1992.

[5] Wolsey LA. Integer programming. vol. 42. Wiley New York; 1998.

[6] Halvorsen R, Mazzoni S, Bryn A, Bakkestuen V. Opportunities for improved distribution modelling practice via a strict maximum likelihood interpretation of MaxEnt. Ecography. 2015;38(2):172–183.

[7] Alex Perkins T, Siraj AS, Ruktanonchai CW, Kraemer MUG, Tatem AJ. Model-based projections of Zika virus infections in childbearing women in the Americas. Nature Microbiology. 2016 jul;1:16126. Available from: http://www.nature.com/articles/nmicrobiol2016126.

[8] Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. eLife [Internet]. 2015 Jun [cited 2016 Apr 19];4:e08347. Available from: http://elifesciences.org/content/4/e08347v3.

[9] 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;10(5):e0004726.

[10] Chouin-Carneiro T, Vega-Rua A, Vazeille M, Yebakima A, Girod R, Goindin D, et al. Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. PLoS Negl Trop Dis. 2016;10(3):e0004543.

[11] Scott TW, Amerasinghe PH, Morrison AC, Lorenz LH, Clark GG, Strickman D, et al. Longitudinal studies of Aedes aegypti (Diptera: Culicidae) in Thailand and Puerto Rico: blood feeding frequency. Journal of medical entomology. 2000;37(1):89–101.

[12] Brady OJ, Johansson MA, Guerra CA, Bhatt S, Golding N, Pigott DM, et al. Modelling adult Aedes aegypti and Aedes albopictus survival at different temperatures in laboratory and field settings. Parasites & Vectors. 2013;6(1):1–12. Available from: http://dx.doi.org/10.1186/1756-3305-6-351.

[13] Chan M, Johansson MA. The incubation periods of dengue viruses. PloS one. 2012;7(11):e50972.

[14] Lessler J, Ott CT, Carcelen AC, Konikoff JM, Williamson J, Bi Q, et al. Times to Key Events in the Course of Zika Infection and their Implications for Surveillance: A Systematic Review and Pooled Analysis. bioRxiv [Internet]. 2016 Mar [cited 2016 Apr 26];Available from: http://biorxiv.org/content/early/2016/03/02/041913.abstract.

[15] USA.com: Texas[Internet]. World Media Group, LCC; 2016 [cited 2016 May 25]. Available from: http://www.usa.com/texas-state.htm.

[16] Wright DH. Correlations Between Incidence and Abundance are Expected by Chance. J Biogeogr [Internet]. 1991 [cited 2016 Apr 19];18(4):463–466. Available from: http://www.jstor.org/stable/2845487.

[17] U S Bureau of Economic Analysis. Texas Counties: Per Capita Income [Internet]; 2014 [cited 2016 Apr 19]. Available from: http://www.txcip.org/tac/census/morecountyinfo.php?MORE=1011.

[18] Duffy MR, Chen TH, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. New Eng J Med [Internet]. 2009 Jun [cited 2016 Apr 4];360(24):2536–2543. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa0805715.

[19] Majumder MS, Cohn E, Fish D, Brownstein JS. Estimating a feasible serial interval range for Zika fever. Bulletin of the World Health Organization [Internet]. 2016 [cited 2016 May 10];Available from: http://www.who.int/bulletin/online_first/16-171009.pdf.

[20] Lloyd AL. Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics. Theoretical Population Biology. 2001;60(1):59–71. Available from: http://www.sciencedirect.com/science/article/pii/S0040580901915254$\delimiter"026E30F$nhttp://www.sciencedirect.com/science/article/pii/S0040580901915254/pdf?md5=26fa247bc639f8948b98845942e70cb3{&}pid=1-s2.0-S0040580901915254-main.pdf.

[21] Centers for Disease Control and Prevention. Draft Interim CDC Zika Response Plan (CONUS and Hawaii): Initial Response to Zika Virus. 2016;Atlanta, Georgia.