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]

Affiliations:

1 - Department 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

Abstract

Background: The southern United States (US) may be vulnerable to outbreaks of Zika Virus (ZIKV), given the broad distribution of ZIKV vector species and periodic ZIKV introductions by travelers returning from affected regions. If autochthonous (locally-acquired) cases appear within the US, policymakers will seek early and accurate indicators of self-sustaining transmission to inform intervention efforts. However, given ZIKV's low reporting rates and the geographic variability in both importations and transmission potential, a small cluster of reported cases may reflect diverse scenarios, ranging from multiple self-limiting but independent introductions to a self-sustaining local epidemic.

Methods and Findings: We developed a stochastic model that captures variation and uncertainty in ZIKV case reporting, importations, and transmission, and applied it to assess county-level risk throughout the state of Texas. For each of the 254 counties, we estimated the future epidemic risk as a function of reported autochthonous cases and evaluated a national recommendation to trigger interventions immediately following the first two reported cases of locally-transmitted ZIKV. Our analysis suggests that the regions of greatest risk for sustained ZIKV transmission include 21 Texas counties along the Texas-Mexico border, in the Houston Metro Area, and throughout the I-35 Corridor from San Antonio to Waco. Variation in vector habitat suitability drives epidemic risk variation, and can be exacerbated by uncertainty in reporting rate. Upon detection of a second locally transmitted case, the threat of epidemic expansion will depend critically on local vulnerability. For high risk Texas counties, we estimate this likelihood to be 64%, assuming an August 2016 risk projection and a 20% reporting rate.

Conclusions: With reliable estimates of key epidemiological parameters, including reporting rates and vector abundance, this framework can help optimize the timing and spatial allocation of public health resources to fight ZIKV in the US.

Introduction

In February 2016, Zika virus (ZIKV) was declared a Public Health Emergency of International Concern [1]. As of 25 July of 2016, the World Health Organization (WHO) confirmed mosquito-transmitted cases in 65 countries and territories, with over 500,000 suspected and confirmed cases in the Americas alone [2,3]. In the US, the potential range of one of the primary vectors for ZIKV, *Aedes aegypti*, is thought to include at least 30 states [4]. Texas ranks among the most vulnerable states for ZIKV transmission, given its suitable climate, international airports, and geographical proximity to affected countries [4–9]. Of the 1,306 imported ZIKV cases in the US, 73 have occurred in Texas. While these importations have yet to spark autochthonous (local) transmission, Texas has historically sustained several small, autochthonous outbreaks (ranging from 4 - 25 confirmed cases) of another arbovirus vectored by *Ae. Aegypti*—dengue (DENV) [10,11].

As peak mosquito season approaches in the US and more cases are introduced via international travelers from the Americas, public health decision makers will face considerable uncertainty in gauging the severity of the threat and in effectively initiating interventions, given the large fraction of undetected ZIKV cases (asymptomatic and symptomatic) as well as the shifting economic balance between intervention expenditure and disease burden [12–15]. Depending on the ZIKV symptomatic fraction, reliability and rapidity of diagnostics, importation rate, and transmission rate, the detection of five autochthonous cases in a single Texas county, for example, may indicate a small chain of cases from a single importation, a self-limiting outbreak, or even a large, hidden epidemic underway (Fig 1). These diverging possibilities have historical precedents. In French Polynesia, a handful of suspected ZIKV cases were reported by October 2013; two months later an estimated 14,000-29,000 individuals had been infected [12,13]. By contrast, Dominica had 18 confirmed cases in early 2016 without a subsequent epidemic [14]. To address the uncertainty, the CDC recently issued conservative guidelines for state and local agencies; they recommend initiation of public health responses following local reporting of *two* non-familial autochthonous ZIKV cases [16].

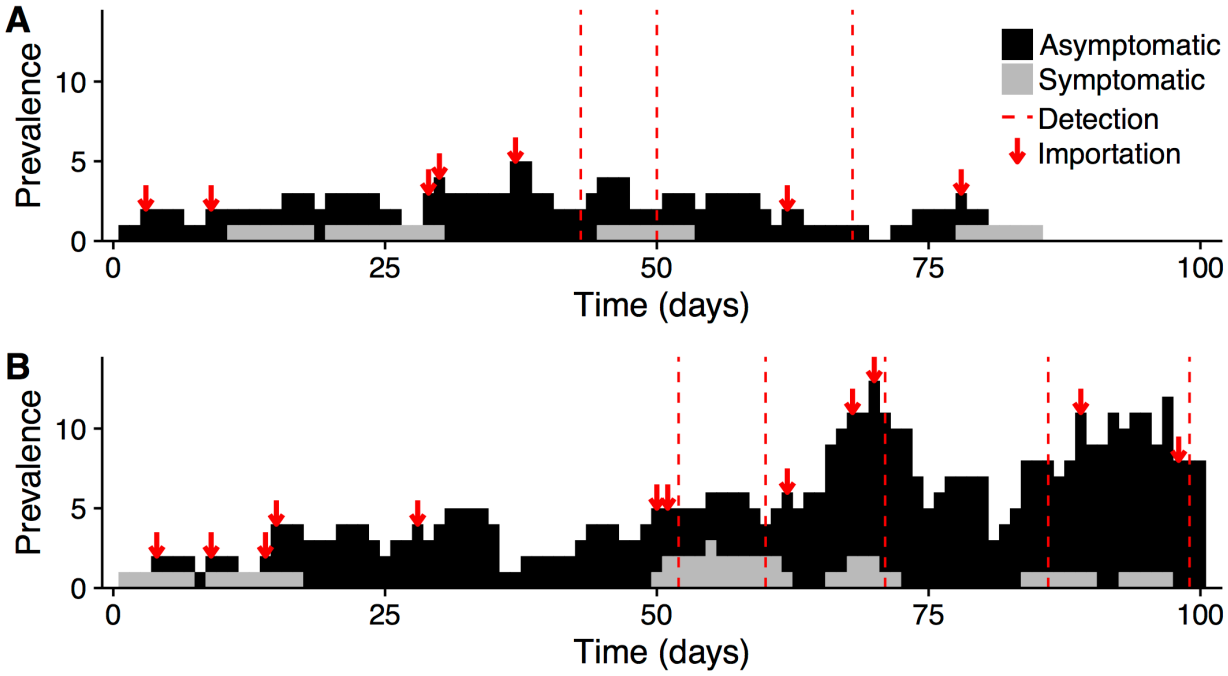


Fig 1. ZIKV emergence scenarios. A ZIKV infection could spark **(A)** a self-limiting outbreak or **(B)** a growing epidemic**.** Cases are partitioned into symptomatic (grey) and asymptomatic (black). Arrows indicate new ZIKV importations by infected travelers and vertical dashed lines indicate case reporting events. On the 75th day, these divergent scenarios are almost indistinguishable to public health surveillance, as exactly three cases have been detected in both. By the 100th day, the outbreak (A) has died out with 21 total infections while the epidemic (B) continues to grow with already 67 total infections. Each scenario is a single stochastic realization of the model with R_0=1.1, reporting rate of 10%, and introduction rate of 0.1 case/day.

Here, we develop a model to support real-time ZIKV risk assessment that accounts for uncertainty regarding ZIKV epidemiology, including importation rates, reporting rates, and local vector population density. This framework can be readily updated as our understanding of ZIKV evolves to provide situational awareness and actionable guidance for public health officials. By simulating ZIKV transmission using a stochastic branching process model [17] based on recent ZIKV data, we estimated ZIKV epidemic risk for each of the 254 counties in Texas, both initially and as a function of the cumulative number of reported autochthonous ZIKV cases. Further, we estimate county-level epidemic risks corresponding to the recommended two-case trigger and demonstrate the design of county-specific triggers indicative of imminent epidemic expansion. Our results suggest that counties along the Texas-

Mexico border, in the Houston Metro Area, and throughout the I-35 Corridor from San Antonio to Waco are at highest risk for sustained epidemics.

Methods

Risk-assessment methodological overview

In our study, we first estimated county-level ZIKV importation rates and sustained transmission risk (R_0) for Texas in August of 2016. For each county, we then used these estimates to simulate 10,000 stochastic ZIKV outbreaks. Finally, we analyzed the simulated outbreaks to determine spatiotemporal variation in ZIKV risk and estimate current and future epidemic risk from the number of reported autochthonous cases (Fig S1).

Estimating County Importation Rates

Our analysis assumes that any ZIKV outbreaks in Texas originate with infected travelers returning from regions with ZIKV activity. To estimate the ZIKV importation rate for specific counties, we (1) estimated the Texas statewide importation rate (expected number of imported cases per day), (2) estimated the probability that the next Texas import will arrive in each county, and (3) took the product of the state importation rate and each county importation probability.

1. During the first quarter of 2016, 27 travel-associated cases of ZIKV were reported in Texas, with 11 occurring in Houston's Harris County [18], yielding a baseline first quarter estimate of 0.3 imported cases per day throughout Texas. In 2014 and 2015, arbovirus introductions into Texas were threefold higher during the third quarter than the first quarter of each year, perhaps driven by seasonal increases in arbovirus activity in endemic regions and the approximately 40% increase from quarter 1 to quarter 3 in total number of international travelers to the US [19]. Taking this as a *baseline* (lower bound) scenario, we projected a corresponding increase in ZIKV importations to 0.9 cases per day (statewide) for the third quarter. We also considered an *elevated* importation scenario, in which the first quarter cases (27) in Texas represent only the

symptomatic (20%) imported cases, corresponding to a projected third quarter statewide importation rate of 4.5 cases per day.

2. We defined county-specific import risk as the probability that the next imported case in Texas will occur in that county. To build a predictive model for import risk, we fit a probabilistic model (maximum entropy) [20] of importation risk to 183 DENV, 38 CHIKV, and 31 ZIKV Texas importations reported at the county level from 2002 to 2016 and informative socioeconomic, environmental, and travel variables (Supplement §1.1). Given the geographic and biological overlap between ZIKV, DENV and Chikungunya (CHIKV), we used historical DENV and CHIKV importation data to supplement ZIKV importations in the importation risk model, while recognizing that future ZIKV importations may be fueled by large epidemic waves in neighboring regions and summer travel, and thus far exceed recent DENV and CHIKV importations [21]. 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. In contrast, ZIKV is not yet as widespread in Mexico as it is in Central and South America, with only one reported ZIKV importation along the border to date (Val Verde County). We included DENV and CHIKV importation data in the model fitting so as to consider potential future importations pressure from Mexico, as ZIKV continues its northward expansion. To find informative predictors for ZIKV importation risk, we analyzed 72 socio-economic, environmental, and travel variables, and removed near duplicate variables and those that contributed least to model performance, based on out-of-sample cross validation [22,23], reducing the original set of 72 variables to 10 (Tables S3-S4).

Estimating County Transmission Risk

The risk of ZIKV emergence following an imported case will depend on the likelihood of mosquito-borne local transmission. For each Texas county, we used the Ross-Macdonald formulation to

148 estimate the ZIKV reproduction number (R_0), which is the average number of secondary infections caused
149 by the introduction of a single infectious individual into a fully susceptible population (Supplement §1.2)
150 [24]. To parameterize the model (Table S5), we used mosquito life history estimates from a combination
151 of DENV and ZIKV studies and estimated *Ae. aegypti* abundance for each county [5]. For parameters that
152 are sensitive to temperature (i.e., mosquito mortality and the extrinsic incubation period), we adjusted the
153 mosquito parameters using average reported Texas county temperatures for the month of August [25].
154
155 ZIKV Outbreak Simulation Model
156 To transmit ZIKV, a mosquito must bite an infected human, the mosquito must get infected with
157 the virus, and then the infected mosquito must bite a susceptible human. We assumed that mosquito-borne
158 transmission would be the main driver of epidemic dynamics, so we did not include sexual transmission
159 in our model. Rather than explicitly model the full transmission cycle, we aggregated the two-part cycle
160 of ZIKV transmission (mosquito-to-human and human-to-mosquito) into a single exposure period, and do
161 not explicitly model mosquitos. For the purposes of this study, we need only ensure that the model
162 produces a realistic human-to-human generation time of ZIKV transmission.
163 The resultant model thus follows a Susceptible-Exposed-Infectious-Recovered (SEIR)
164 transmission process stemming from a single ZIKV infection using a Markov branching process model
165 (Fig S2). The temporal evolution of the compartments is governed by daily probabilities of infected
166 individuals transitioning between S, E, I, and R states, and new ZIKV cases arising from importations or
167 autochthonous transmission (Table S6). We treat days as discrete generations, and the next disease state
168 progression depends solely on the current state and the transition probabilities. We assume that infectious
169 cases cause a Poisson distributed number of secondary cases per day (via human to mosquito to human
170 transmission), and infectious individuals are introduced daily according to a Poisson distributed number
171 of cases around the importation. Furthermore, *Infectious* cases are categorized into reported and
172 unreported cases according to a reporting rate. We assume that reporting rates approximately correspond
173 to the percentage (~20%) of symptomatic ZIKV infections [14] and occur at the same rate for imported

and locally acquired cases. Additionally we make the simplifying assumption that reported cases transmit ZIKV at the same rate as unreported cases. We track imported and autochthonous cases separately, and conduct risk analyses based on reported autochthonous cases only, under the assumption that public health officials will have immediate and reliable travel histories for all reported cases [16].

To accurately model the timing of ZIKV outbreaks, we fit the ZIKV generation time to recent estimates (Supplement §2.4) [26]. 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 [26]. In our model, the generation time corresponds to the exposure period followed by half of the infectious period. First, we fit the infectious period in our model to human ZIKV estimates for duration of viral shedding, which we assumed to be the length of the infectious period but may be an underestimate of the total length. Specifically, we solved for transition rates of a Boxcar Model [27] that produce an infectious period with mean duration of 9.88 days (Table S6) [28]. Then, we fit the exposure period to achieve the empirical ZIKV generation time distribution [26], yielding a mean exposure period of 10.4 days (95% CI 6-17) and 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 [28]. While there is much uncertainty regarding ZIKV transmission, our model framework is flexible to extensions and updates as we learn more about ZIKV epidemiology including reporting rates, transmission probabilities, clinical features, and outside importation pressures.

Simulations

For each county risk scenario, defined by a specified importation rate, transmission rate, and reporting rate, we ran 10,000 stochastic simulations. Each simulation began with a single imported infectious case and terminated either when there were no individuals left 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 all simulations that fulfill two

criteria: reach 2,000 cumulative autochthonous infections and have a maximum daily prevalence exceeding 50 autochthonous cases (Figs S3).

Outbreak Analysis

Policymakers must often make decisions in the face of uncertainty, such as when and where to initiate ZIKV interventions. Our stochastic framework allows us to provide real-time county-level risk assessments as reported cases accumulate. For each county, we found the probability that an outbreak will progress into an epidemic (reach 2,000 cases with a maximum daily prevalence over 50), as a function of the number of reported cases. We call this *epidemic risk*. To solve for epidemic risk in a county following the *x*th reported autochthonous case, we first find all simulations (of 10,000 total) that experience at least *x* reported autochthonous cases, and then calculate the proportion of those that are ultimately classified as epidemics. For example, consider a county in which 1,000 of 10,000 simulated outbreaks reach at least two reported autochthonous cases; the remaining 9,000 simulations dissipate with only one or zero case reports. If only 50 of the 1,000 simulations ultimately fulfill the two epidemic criteria, then the estimated epidemic risk following two reported cases in that county would be 5%. This simple classification scheme performs quite well, only rarely misclassifying a string of small outbreaks as an epidemic, with the probability of such an error increasing with the importation rate. For example, epidemics should not occur when R_0=0.9. If the importation rate is high, however, overlapping series of moderate outbreaks may occasionally meet the two epidemic criteria. Even under the highest importation rate we considered (0.3 cases/day), only 1% of outbreaks were misclassified.

This method can be applied to evaluate universal triggers (like the recently recommended two-case trigger) or derive robust triggers based on local importation and transmission risks as well as the risk tolerance of public health agencies. For example, if a policymaker would like to initiate interventions as soon as the risk of an epidemic reaches 30%, we would simulate local ZIKV transmission and solve for the number of reported cases at which the probability of an epidemic first exceeds 30%. Generally, the recommended triggers decrease as the policymaker threshold for action decreases (for example,

policymakers would act sooner (fewer reported cases) for a 10% versus 30% threshold) and as the local transmission potential increases (*e.g.* R_0 = 1.5 versus R_0 = 1.2). A policymaker wishing to trigger interventions early, upon even a low probability of epidemic spread, has a low tolerance for failing to intervene but may waste resources on unnecessary interventions; a policymaker willing to wait longer, has a higher risk tolerance, but may implement interventions too late in the course of the outbreak.

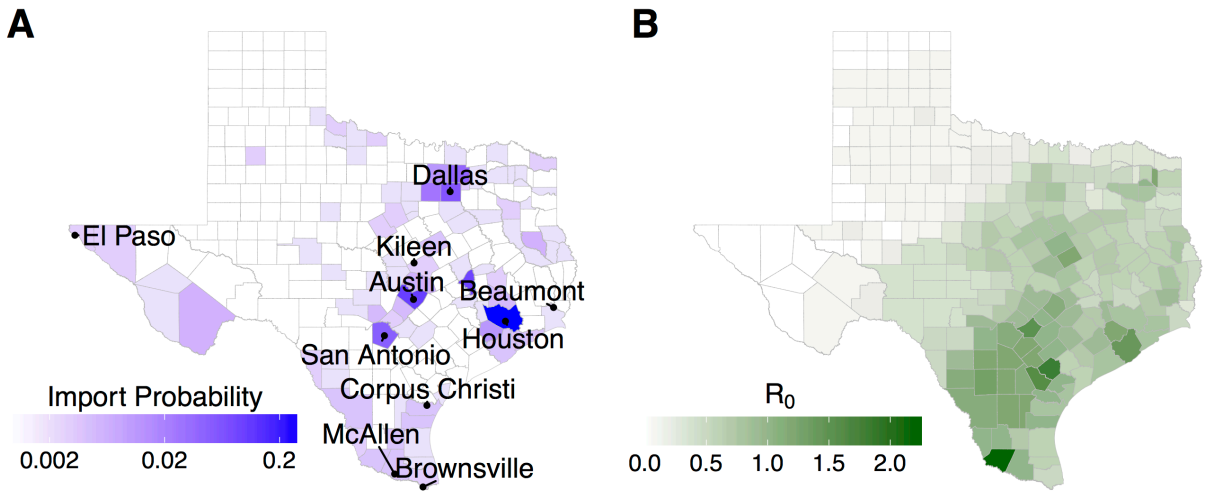
County Uncertainty Analysis

We took two approaches to addressing uncertainty in the model parameters. First, we conducted a sensitivity analysis to address the considerable uncertainty regarding several inputs into our estimation of R_0, including mosquito biology, ZIKV epidemiology, and human-mosquito interactions (Supplement §4). For most factors, the county estimates of R_0 simply scale linearly with changes in the factor. However, county-specific human-mosquito contact rates can change relative county risks based on assumptions regarding the socioeconomic effect on human-mosquito interactions (Fig S5-6), and county risk moves southward as the summer heat subsides (Fig S7). Second, given the considerable uncertainty regarding ZIKV epidemiology, we examined a scenario where the absolute values of both R_0 and importation rate are unknown, but lie within plausible ranges for Texas. To do so, we randomly sampled 10,000 simulations from the high risk Texas county outbreaks (counties with R_0>1), creating an amalgamous high risk county, and completed the outbreak analysis as we do with individual counties.

Results

To develop a ZIKV risk assessment framework for Texas counties, we first estimate county-level ZIKV importation and transmission rates for August 2016. ZIKV importation risk within Texas is predicted by variables reflecting urbanization, mobility patterns, and socioeconomic status (Table S3), and is concentrated in metropolitan counties of Texas (Fig 2A). The two highest risk counties--Harris, which includes Houston and has an estimated 27% chance of receiving the next imported Texas case, and Travis, which includes Austin and has a 10% chance of the next importation--contain international

252 airports. Other high risk regions include Brazos County, the Dallas and San Antonio metropolitan areas,
253 and several counties along the Texas-Mexico border.
254



255

256 **Fig 2. ZIKV importation and transmission risk estimates across Texas. (A)** Color indicates the probability that the next
257 ZIKV import will occur in a given county for each of the 254 Texas counties. Probability is colored on a log scale. The 10 most
258 populous cities in Texas are labeled. Houston's Harris County has 2.7 times greater chance than Austin's Travis County of
259 receiving the next imported case. **(B)** Estimated county-level R_0 for ZIKV (See Fig S5-11 for sensitivity analysis).
260

261 Our county-level estimates of autochthonous ZIKV transmission risk (Fig 2B) suggest that the
262 majority of Texas counties (87%) have an estimated R_0 below one, and thus are unlikely to sustain
263 epidemics. The Southeast region of Texas has the highest estimated transmission risk, driven primarily by
264 high mosquito habitat suitability. These estimates are sensitive to uncertainty in several parameters (Fig
265 S5-11), and can be updated as we learn more about ZIKV. While the average transmission risk may be
266 higher or lower than our baseline assumption, and will certainly vary seasonally, the relative risks of
267 regions and counties are robust (Fig S6-S10), and allow us to conduct plausible case studies and identify
268 at risk areas for enhanced surveillance and preparedness efforts. Given the uncertainty underlying the

county R_0 estimates, we also aggregate the 21 highest estimates into a plausible distribution for a high risk Texas county, ranging from $R_0 = 1.0$ to $R_0 = 2.2$ with a median of 1.1.

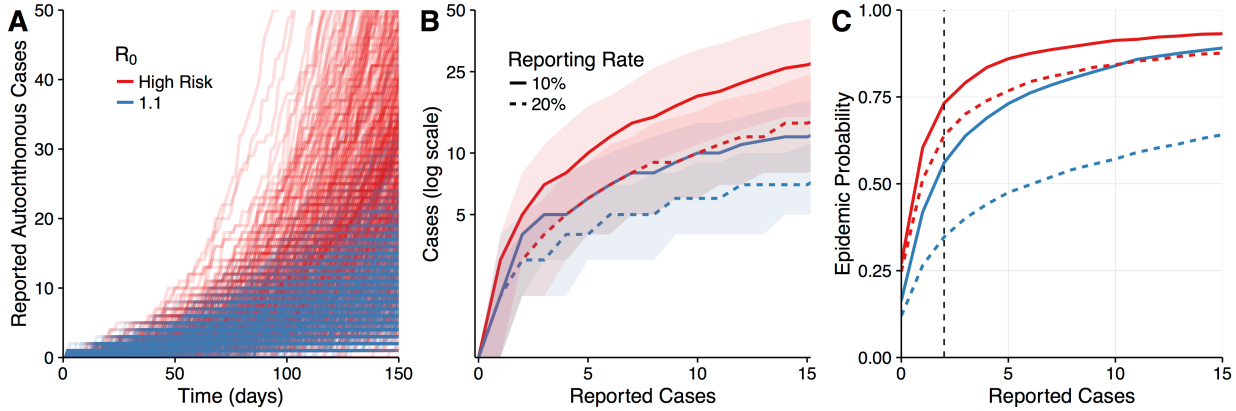


Fig 3. Real-time risk-assessment for ZIKV situational awareness and forecasting. (A) Simulated outbreaks, assuming an importation rate of 0.01 case per day, for a known (moderate risk) R_0 (blue) or an unknown high risk R_0 (red). Two thousand randomly sampled simulations are shown for each scenario. (B) Total number of (current) autochthonous cases as a function of the cumulative reported autochthonous cases, assuming an importation rate of 0.01 case per day, for a known R_0 (blue) or an unknown high-risk R_0 (red), and a relatively high (dashed) or low (solid) reporting rate. Ribbons indicate 50% quantiles. (C) The increasing probability of imminent epidemic expansion as reported autochthonous cases accumulate, compared across the high risk (red) and known moderate risk (blue) for a low (solid) and high (dashed) reporting rate. Suppose cases arise in a high risk county and a policymaker wishes to trigger a response as soon as two cases are reported (vertical line). With a 20% reporting rate there is a 64% probability of an ensuing epidemic.

Under a single set of epidemiological conditions, wide ranges of outbreaks are possible (Fig 3A). The relationship between what policymakers can observe (cumulative reported cases) and what they wish to know (current prevalence) can be obscured by such uncertainty, and will depend critically on both the transmission and reporting rates (Fig 3B). If key drivers, such as R_0, can be estimated with confidence, then the breadth of possibilities narrows, enabling more precise surveillance. For example, under a known moderate R_0 scenario and with a 20% reporting rate, ten cumulative reported cases corresponds to an expected prevalence of 6 cases with a 95% CI of 1-16; under an unknown but high R_0 scenario, the same

number of cases corresponds to an expected prevalence of 10 cases with a much wider 95% CI of 2-32 (Fig 3B).

We apply our model to characterize time-varying epidemic risk as cases accumulate in a given county. Under both a known moderate risk and unknown high risk scenario, we track the probability of epidemic expansion following each additional reported case (Fig 3C). Across the full range of reported cases, the probability of epidemic spread is always higher in the high risk scenario, with the moderate risk scenario showing more sensitivity to the reporting rate. These curves can support both real-time risk assessment as cases accumulate and the identification of surveillance triggers indicating when risk exceeds a specified threshold. For example, suppose a policymaker wanted to initiate an intervention upon two reported cases and thought the reporting rate was 20%, this would correspond with a 64% probability of an epidemic in the high risk county but only 35% in the known moderate risk county. Alternatively suppose a policy maker wishes to initiate an intervention when the chance of an epidemic exceeds 50%. In the high risk scenario, they should act immediately following the 1st reported case; in the moderate risk scenario, the corresponding trigger ranges from two to seven reported cases, depending on the reporting rate. As the policymaker's threshold (risk tolerance) increases, the recommended surveillance triggers can be adjusted accordingly.

To evaluate a universal intervention trigger of two reported autochthonous cases, we estimate both the probability of a trigger event (two such cases) in each county and the level of epidemic risk at the moment a trigger event occurs (second case reported) in each county. Assuming a baseline importation rate extrapolated from recent importations to August 2016 and a 20% reporting rate, only a minority of counties are likely to experience a trigger event (Fig 4A). While 231 of the 254 counties (91%) have non-zero probabilities of experiencing two reported autochthonous cases, only 63 counties have at least a 10% chance of such an event, with the remaining 168 counties having a median probability of 0.017 (range 0.0004 to 0.089). Next, assuming that a second autochthonous case has indeed been reported, we find that the underlying epidemic risk varies widely, with most counties having near zero epidemic probabilities and a few counties far exceeding a 50% chance of epidemic expansion. For example, two reported

autochthonous cases in Starr County, along the Texas-Mexico border, correspond to a 99% chance of ongoing transmission. The greater San Antonio metropolitan region appears to be the highest risk metropolitan region with four of its eight counties having a higher than 25% probability of experiencing two reported autochthonous cases; in those four counties, the epidemic risk upon detection of a second case ranges from 19-90%. Houston metropolitan region is also a high risk region with its second (Fort Bend) and fourth (Brazoria) largest counties having a 39% and 45% chance of sustaining two reported autochthonous cases, respectively, with corresponding epidemic risks of 67% and 86% thereafter.

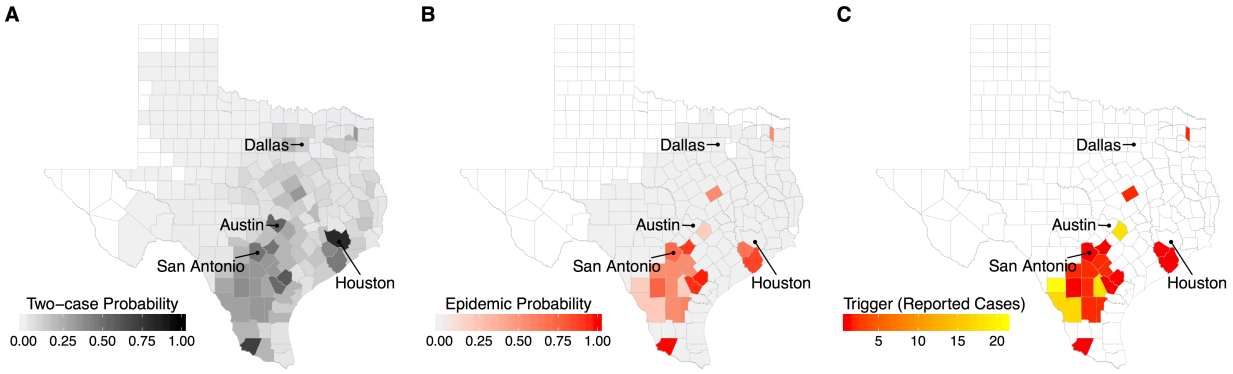


Fig 4. Texas county ZIKV risk assessment. (A) Probability of an outbreak with at least two reported autochthonous ZIKV cases. **(B)** The probability of epidemic expansion at the moment the second autochthonous ZIKV case is reported in a county. White counties never reach two reported cases across all 10,000 simulated outbreaks; light gray counties reach two cases, but never experience epidemics. **(C)** Recommended county-level surveillance triggers (number of reported autochthonous cases) indicating that the probability of epidemic expansion has exceeded 50%. White counties indicate that fewer than 1% of the 10,000 simulated outbreaks reached two reported cases. All three maps assume a 20% reporting rate and a baseline importation scenario for August 2016 (81 cases statewide per 90 days) projected from historical arbovirus data. (Figure S4 provides corresponding estimates under a worse case elevated importation scenario).

Given that a universal trigger may signal highly disparate levels of ZIKV risk, policymakers might seek to adapt their triggers to local conditions. Suppose a policymaker wishes to design trigger that indicate a 50% chance of an emerging epidemic (Fig 4C). Under the baseline importation and reporting rates, only 21 of the 254 counties in Texas are expected to reach a 50% epidemic probability, with

triggers ranging from one (Starr County) to 21 (Dimmit County) reported autochthonous cases, with a median of two cases. The remaining counties have less than a 1% chance of experiencing sustained ZIKV transmission. Under an elevated importation scenario, assuming that only one fifth of ZIKV importations (the symptomatic proportion) have been observed, we find that the recommended triggers decrease by a mean of 1.5 reported cases (Fig S4) and the size of Texas' population at risk for sustained ZIKV transmission is expected to increase from ~14% to ~30%, largely driven by increased risk in the Houston metropolitan area.

Discussion

US public health authorities are responding to ZIKV importations and preparing for the possibility of ZIKV outbreaks in vulnerable regions this season and in future seasons. A key challenge is knowing when and where to initiate interventions based on potentially sparse and biased ZIKV case reports. Our simple model is designed to address this challenge by providing county-specific epidemic risk estimates as a function of reported cases. We demonstrate its application across the 254 ecologically and demographically diverse counties of Texas, a high risk state [6,7,9]. Based on county-level estimates for ZIKV importation and transmission rates (Fig 2), we expect that most Texas counties are not at risk for a sustained ZIKV epidemic (Fig 4). However, 30% of Texas' population may reside in vulnerable regions, including the cities of Austin, San Antonio, and Waco along the I-35 corridor, Houston, and the Rio Grande Valley. The higher the ZIKV importation rate in these locations, the higher the chance of an epidemic (Fig S4). However, even in the most high risk regions of Texas, we expect far more limited ZIKV transmission than observed in Central America, South America, and Puerto Rico, where R_0 has been estimated to be as high as 11 [24,29,30]. Our analysis is consistent with recent introductions of DENV and CHIKV into Texas that have failed to spark large epidemics.

Surveillance triggers--guidelines specifying situations that warrant intervention--are a key component of many public health response plans. Given the urgency and uncertainty surrounding ZIKV, *universal* recommendations can be both pragmatic and judicious. In choosing an appropriate trigger for an

environmentally and socioeconomically diverse region, policymakers must weigh the risks of acting unnecessarily with those of responding too late. A risk averse policymaker will likely select a low trigger (a small number of reported cases), ensuring early detection in the riskiest sites at the cost of false alarms in low risk locations (Fig 4C). In that vein, the CDC recently issued conservative guidelines to state and local public health agencies that suggest a two-case trigger for initiating interventions [16].

To assist Texas policymakers in interpreting this recommendation, we assessed the likelihood and implication of a two-case trigger for each of Texas' 254 counties, under a scenario projected from recent ZIKV data to August 2016. Across counties, there is enormous variation in both the chance of a trigger and the magnitude of the public health threat if and when two cases are reported. If and when two autochthonous ZIKV cases are reported, only 18 of Texas' 254 counties have over a 20% chance of experiencing sustained transmission (under our baseline scenario); in most of the remaining 236 counties, the threat is much lower.

Rather than implement a universal *trigger,* which may indicate different threats in different locations, one could design local surveillance triggers that correspond to a universal *risk threshold*. Our modeling framework can readily identify triggers (numbers of reported cases) for indicating any specified epidemic event (e.g., prevalence reaching a threshold or imminent epidemic expansion) with any specified risk tolerance (e.g., 10% or 50% chance of that the event has occurred), given local epidemiological conditions. As a case study, we identify *epidemic expansion* triggers in each of Texas' 254 counties, each designed to indicate when the probability of an epidemic exceeds 50%. Across the 21 counties with non-negligible probabilities of an epidemic, the recommended triggers range from one to 21 reported autochthonous cases, highlighting Texas' spatial risk heterogeneity. These findings apply only to the early, pre-epidemic phase of ZIKV in Texas, when travel from affected regions outside the contiguous US is the primary importation source. If self-sustaining outbreaks emerge within Texas, there may be county-to-county importations, particularly in high risk regions, that are not yet included in the model.

This simple framework offers a flexible means for bringing current data and expert knowledge to assist critical public health decision making. The design of a trigger--both the event to be detected and the

probability threshold upon which to take action--requires extensive public health expertise and deliberation. Our case studies were motivated by formal recommendations and informal discussions with state and national public health agencies. They demonstrate our data-driven approach for relating observed disease activity to underlying risk in the face of great uncertainty, and address the pros, cons and complexity of developing universal versus regional action plans. However, they are not meant to critique or validate any specific policies, and highlight the prudence of recent ZIKV planning efforts.

Importantly, our analyses rest on the recent and limited scientific investigations of ZIKV's biology and epidemiology, and should be continually updated as our understanding matures. Specifically, our county-level estimates of R_0 are sensitive to many underlying assumptions, particularly regarding temperature-dependent and socioeconomic factors (Supplement §4). Although the risk of ZIKV transmission may be much higher or lower than assumed in our baseline scenario, the relative vulnerabilities of counties are fairly robust to our assumptions. Thus, counties with the highest estimated risks should likely be prioritized for surveillance and interventions resources. Given the minimal incursions of DENV and CHIKV into Texas, we suspect that, if anything, we may be underestimating the socioeconomic and behavioral impediments to ZIKV transmission in the contiguous US, and thus overestimating the transmission risk across Texas. While it is beyond the scope of the current paper, these estimates may be further refined as more detailed arbovirus importation and outbreak data become available.

The reporting rate dictates the relationship between observed cases and the underlying outbreak, and its magnitude impacts the timeliness and precision of detection. If only a small fraction of cases are reported, the first few reported cases may correspond to a wide range of underlying epidemiological conditions, from isolated introductions to a growing epidemic. In contrast, if most cases are reported, policymakers can wait longer (in terms of the number of reported cases) to trigger interventions and have more confidence in their epidemiological assessments. ZIKV reporting rates are expected to remain quite low, because an estimated 80% of infections are asymptomatic, and DENV reporting rates have historically matched its asymptomatic proportion [14,31]. Obtaining a realistic estimate of the ZIKV

reporting rate is arguably as important as increasing the rate itself, with respect to reliable situational awareness and forecasting. An estimated 8-22% of ZIKV infections were reported during the 2013-2014 outbreak in French Polynesia [30]; similarly, an estimated 10% have been reported during the ongoing epidemic in Columbia [29]. While these provide a baseline estimate for the US, there are many factors that could increase (or decrease) the reporting rate, such as ZIKV awareness among both the public and health-care practitioners. Thus, rapid estimation of the reporting rate should be a high priority. While some methods require extensive epidemiological data not typically available early in an outbreak [32], a new method exploiting early outbreak viral sequence data was introduced during the recent West African Ebola epidemic [33]. However, as of July 2016, there are no US ZIKV sequences available on GenBank and few available from other regions.

Unlike other estimation methods, this framework is designed to support risk assessments both prior to ZIKV outbreaks and, in real-time, upon the early detection of cases through the lens of potentially sparse and biased surveillance data [34,35]. It can facilitate situational awareness, allowing analysts to translate case counts into estimates of overall prevalence and the potential for future spread, albeit with large error bars. It can also support the development of response plans, by forcing policymakers to be explicit about risk tolerance, that is, the certainty needed before sounding an alarm, and quantifying the consequences of premature or delayed interventions. For example, should ZIKV-related pregnancy advisories be issued when there is only 5% chance of an impending epidemic? 10% chance? 80%? A policymaker has to weigh the costs of false positives--resulting in unnecessary fear and/or intervention--and false negatives--resulting in suboptimal disease control and prevention--complicated by the difficulty inherent in distinguishing a false positive from a successful intervention. The more risk averse the policymaker (with respect to false negatives), the earlier the trigger should be, which can be exacerbated by low reporting rates, high importation rate, and inherent ZIKV transmission potential. In ZIKV prone regions with low reporting rates, even risk tolerant policymakers should act quickly upon seeing initial cases; in lower risk regions, longer waiting periods may be prudent.

This approach can be readily applied to other at risk states in the US. Given that Florida has experienced a higher importation of ZIKV cases than Texas (more than 270 as of the end of July 2016) and more locally acquired DENV cases since 2009 [36–38], the probability of sustained ZIKV transmission and public health threat following a second reported autochthonous cases may be higher in Florida than estimated for Texas.

Acknowledgements

We acknowledge the Texas Department of State Health Services (DSHS) for providing historic arbovirus county importation data, and MUG Kraemer and A Perkins for providing Texas mosquito abundance data and technical guidance. We also thank M Johansson and M Meltzer of the Centers for Disease Control and Prevention (CDC) and Dr. John Hellerstedt, commissioner of the Texas DSHS, for critical conversations regarding the modeling framework and its application to Texas and national ZIKV planning efforts. We finally acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing High performance computing resources that have contributed to the research results reported within this paper. URL: http://www.tacc.utexas.edu.

Code Availability

The R package, rtZIKVrisk, has been created to accompany the manuscript and can be found here: https://github.com/sjfox/rtZIKVrisk. The package provides functionality to simulate and analyze ZIKV outbreaks according to our framework, and recreate the figures presented in the manuscript.

References

1. Gulland A. Zika virus is a global public health emergency, declares WHO. BMJ. BMJ Publishing Group Ltd; 2016;352. doi:10.1136/bmj.i657

2. PAHO, WHO. Suspected and confirmed Zika cases reported by countries and territories in the Americas, 2015-2016 [Internet]. 2016 p. http://ais.paho.org/phip/viz/ed_zika_epicurve.asp.

3. WHO. Situation Report: Zika Virus, Microcephaly, and Guillan-Barre Syndrome 14 July 2016. 2016.

4. About estimated range of Aedes aegypti and Aedes albopictus in the United States, 2016 Maps. In: Atlanta: Centers for Disease Control and Prevention [Internet]. [cited 1 Jan 2016]. Available: http://www.cdc.gov/zika/vector/range.html

5. 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. eLife Sciences Publications Limited; 2015;4: e08347. Available: http://elifesciences.org/content/4/e08347v3

6. Messina JP, Kraemer MU, Brady OJ, Pigott DM, Shearer FM, Weiss DJ, et al. Mapping global environmental suitability for Zika virus. Elife. eLife Sciences Publications Limited; 2016;5: e15272. Available: https://elifesciences.org/content/5/e15272

7. Gardner LM, Chen N, Sarkar S. Global risk of Zika virus depends critically on vector status of Aedes albopictus. Lancet Infect Dis. Elsevier; 2016;16: 522–523. Available: http://www.thelancet.com/article/S1473309916001766/fulltext

8. Peterson AT, Osorio J, Qiao H, Escobar LE. Zika virus, elevation, and transmission risk. PLoS Curr Outbreaks. 2016;9. doi:10.1371/currents.outbreaks.a832cf06c4bf89fb2e15cb29d374f9de

9. Monaghan AJ, Morin CW, Steinhoff DF, Wilhelmi O, Hayden M, Quattrochi DA, et al. On the Seasonal Occurrence and Abundance of the Zika Virus Vector Mosquito Aedes Aegypti in the Contiguous United States. PLoS Curr. 2016; 1–31. doi:10.1371/currents.outbreaks.50dfc7f46798675fc63e7d7da563da76

10. Texas Department of State Health and Human Services. Arbovirus Activity in Texas 2014 Surveillance Report [Internet]. 2014. Available: 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&s

11. CDC, Texas Department of State Health and Human Services. Dengue Hemorrhagic Fever --- U.S.-Mexico Border, 2005 [Internet]. 2007. Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5631a1.htm

12. Lessler J, Chaisson LH, Kucirka LM, Bi Q, Grantz K, Salje H, et al. Assessing the global threat from Zika virus. Science (80-). 2016;8160. doi:10.1126/science.aaf8160

13. Lloyd-Smith JO, Funk S, McLean AR, Riley S, Wood JLN. Nine challenges in modelling the emergence of novel pathogens. Epidemics. Elsevier B.V.; 2014;10: 35–39. doi:10.1016/j.epidem.2014.09.002

14. Duffy MR, Chen T-H, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika Virus Outbreak on Yap Island, Federated States of Micronesia. N Engl J Med. 2009;360: 2536–2543. doi:10.1056/NEJMoa0805715

15. Alfaro-Murillo JA, Parpia AS, Fitzpatrick MC, Tamagnan JA, Medlock J, Ndeffo-Mbah ML, et al. A Cost-Effectiveness Tool for Informing Policies on Zika Virus Control. Carabin H, editor. PLoS Negl Trop Dis. 2016;10: e0004743. doi:10.1371/journal.pntd.0004743

16. Interim CDC recommendations for Zika vector control in the continental United States. In: Centers for Disease Control and Prevention [Internet]. 2016 [cited 5 Jan 2016]. Available: http://www.cdc.gov/zika/public-health-partners/vector-control-us.html

17. Getz WM, Lloyd-Smith JO. Basic methods for modeling the invasion and spread of contagious diseases. Dis Evol Model concepts, data Anal AMS-DIMACS Ser. 2006;71: 87. Available: http://books.google.com/books?hl=en&lr=&id=5uMf4qCmghEC&oi=fnd&pg=PA87&dq=Basic+Methods+for+Modeling+the+Invasion+and+Spread+of+Contagious+Diseases&ots=M3miy-szrR&sig=Re_k7i7YQoXGHkktCbLilsv41XQ\nhttp://books.google.com/books?h

18. Zika in Texas. In: Texas Department of State Health Services [Internet]. 2016 [cited 5 Apr 2016]. Available: http://www.texaszika.org/

19. Office of Travel & Tourism Industries. US Monthly Arrivals Trend Line [Internet]. 2014.

Available: http://travel.trade.gov/view/m-2014-I-001/index.html

20. Jaynes ET. Information Theory and Statistical Mechanics. Phys Rev. American Physical Society; 1957;106: 620–630. doi:10.1103/PhysRev.106.620

21. Musso D, Cao-Lormeau VM, Gubler DJ. Zika virus: following the path of dengue and chikungunya? Lancet. Elsevier Ltd; 2015;386: 243–244. doi:10.1016/S0140-6736(15)61273-9

22. Wolsey LA. Integer programming. Wiley; 1998.

23. Merow C, Smith MJ, Silander JA. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography (Cop). 2013;36: 1058–1069. doi:10.1111/j.1600-0587.2013.07872.x

24. Alex Perkins T, Siraj AS, Ruktanonchai CW, Kraemer MUG, Tatem AJ. Model-based projections of Zika virus infections in childbearing women in the Americas. Nat Microbiol. 2016;1: 16126. doi:10.1038/nmicrobiol.2016.126

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

26. Majumder MS, Cohn E, Fish D, Brownstein JS. Estimating a feasible serial interval range for Zika fever. Bull World Health Organ. 2016; doi:10.2471/BLT.16.171009

27. Lloyd AL. Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics. Theor Popul Biol. 2001;60: 59–71. doi:10.1006/tpbi.2001.1525

28. Lessler JT, 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: a systematic review and pooled analysis. Bull World Health Organ. 2016; doi:10.2471/BLT.16.174540

29. Rojas DP, Dean NE, Yang Y, Kenah E, Quintero J, Tomasi S, et al. The Epidemiology and Transmissibility of Zika Virus in Girardot and San Andres Island, Colombia. bioRxiv. 2016; Available: http://biorxiv.org/content/early/2016/04/24/049957.abstract

30. Kucharski AJ, Funk S, Eggo RM, Mallet H, Edmunds WJ, Nilles EJ. Transmission dynamics of Zika virus in island populations : a modelling analysis of the 2013 – 14 French Polynesia outbreak.

543 bioRxiv. 2016; 1–15.

544 31. Dechant E, Rigau-Perez J. Hospitalizations for suspected dengue in Puerto Rico, 1991-1995:
545 estimation by capture-recapture methods. The Puerto Rico Association of Epidemiologists. Am J
546 Trop Med Hyg. 1999;61: 574–578. Available: http://www.ajtmh.org/content/61/4/574.short

547 32. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution
548 and burden of dengue. Nature. Nature Publishing Group; 2013;496: 504–507.
549 doi:10.1038/nature12060

550 33. Scarpino S V, Iamarino A, Wells C, Yamin D, Ndeffo-Mbah M, Wenzel NS, et al.
551 Epidemiological and viral genomic sequence analysis of the 2014 ebola outbreak reveals clustered
552 transmission. Clin Infect Dis. Oxford University Press; 2015;60: 1079–82. Available:
553 http://cid.oxfordjournals.org/content/early/2015/01/05/cid.ciu1131.full

554 34. Cauchemez S, Epperson S, Biggerstaff M, Swerdlow D, Finelli L, Ferguson NM. Using Routine
555 Surveillance Data to Estimate the Epidemic Potential of Emerging Zoonoses: Application to the
556 Emergence of US Swine Origin Influenza A H3N2v Virus. Peiris JSM, editor. PLoS Med.
557 2013;10: e1001399. doi:10.1371/journal.pmed.1001399

558 35. Blumberg S, Lloyd-Smith JO. Comparing methods for estimating R0 from the size distribution of
559 subcritical transmission chains. Epidemics. 2013;5: 131–145. doi:10.1016/j.epidem.2013.05.002

560 36. Teets FD, Ramgopal MN, Sweeney KD, Graham AS, Michael SF, Isern S. Origin of the dengue
561 virus outbreak in Martin County, Florida, USA 2013. Virol Reports. NIH Public Access; 2014;1-2:
562 2–8. doi:10.1016/j.virep.2014.05.001

563 37. Rey J. Dengue in Florida (USA). Insects. 2014;5: 991–1000. doi:10.3390/insects5040991

564 38. Centers for Disease Control and Prevention. Zika virus disease in the United States, 2015–2016.
565 Available: https://www.cdc.gov/zika/geo/united-states.html

566

567

568 **S1 Supporting Information.** Supporting information for "Real-time Zika risk assessment in the
569 United States."

570 **S2 Supporting Data.** Raw Texas county arbovirus importation data used for estimating county ZIKV
571 importation risk. Data provided by Texas DSHS.

572 **S3 Supporting Data.** Raw county socioeconomic data used for estimating county ZIKV importation risk.
573 Data accessed from: https://www.census.gov/programs-surveys/acs/data.html

574 **S4 Supporting Data.** Texas county ZIKV risk estimates used as inputs for the ZIKV simulation model.
575 Outputs of the importation and transmission risk analysis.

576