**The Dissertation Committee for Spencer James Fox
certifies that this is the approved version of the following dissertation:**

Statistical modeling of disease emergence

Committee:

Lauren Ancel Meyers, Supervisor

Claus O. Wilke

James J. Bull

Shweta Bansal

Michael J. Daniels

Statistical modeling of disease emergence

by

Spencer James Fox

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2018

To my friends and family

Acknowledgments

This dissertation would not have happened without the love and support of my family. To my parents, Brad and Kristine Fox: I learned my first programming language (Logo) at one of those summer science workshops you forced me to go to when I was a kid, and I am forever grateful for that. You've always known the right amount to push me to improve myself, and have been there to support me if and when I failed. As for my brother, Ben Fox: you've been a continuous reminder that there is a life outside of school, and an aspirational image for what that life can look like. Thank you for always being there for me.

I want to thank my advisor, Dr. Lauren Ancel Meyers, for her unending mentorship and support. She allowed me the freedom to pursue the projects that caught my interest, and helped me find focus when I inevitably found my way into a rabbit hole. Steve Bellan served the role of an unofficial secondary advisor to me as a post-doc in the lab, and I'm grateful for his advice and friendship. The sixth floor hallway hasn't been the same since he left. I also want to thank my other lab members for their input on numerous presentations and harebrained ideas.

The work presented in this thesis was drastically improved by the input of a number of people and I would like to thank them for their time and

effort. My committee members: Dr. Claus Wilke, Dr. Shweta Bansal, Dr. Michael Daniels, and Dr. James Bull, whose guidance has greatly expedited and improved my work. Second, my collaborators: Lauren Castro, Xi Chen, Kai Liu, Steve Bellan, Ned Dmitrov, Alison Galvani, Joel Miller, Alex Perkins, and Michael Johansson, whose expertise and input were imperative to the completion of my work.

More indirectly, this dissertation would never have been produced without a long list of important teachers. I'm thankful to Mrs. Paradise and Mrs. Wagner who first introduced me to science and critical thinking. My interest in biology dates back to Mrs. Gleicher's classes in high school, when she singlehandedly prepared me for college biology. I'm thankful to Dr. Gabi Waite who advised my first summer research experience where I first became interested in pursuing a PhD, and to Dr. Richard Anthony for his continued mentorship even after the completion of my senior thesis research project. Dr. Peter Coppinger first showed me that science can also be fun, as he taught us how to brew beer well before it was cool to brew beer. Finally, to Dr. Ella Ingram: I came for your required classes, but stayed for your collaboration and friendship. There is no one who was more directly important to my decision to apply to EEB programs and to my interest in infectious diseases.

My acknowledgements could not be complete without mentioning the number of people maintaining and producing open source software and materials. The R programming language has been the single most important tool I've used in graduate school, and I am grateful to the R Core team and RStudio for

Finally, I would like to thank Anne Chambers. Over the past three years you have probably put up with more conversations about R_0 than anyone deserves in a whole lifetime and you haven't complained once. Thank you for making me a better scientist, naturalist, outdoorsman, and person, and most importantly for bringing continual enjoyment to my life.

Statistical modeling of disease emergence

Spencer James Fox, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Lauren Ancel Meyers

Infectious diseases seem to be appearing at an unprecedented rate: within the last few years alone, a sequence of novel diseases like MERS-CoV, Chikungunya, and Zika have emerged. Concurrently, a number of previously known diseases have re-emerged like the 2009 H1N1 pandemic and the 2014 Ebola epidemic. While these known and unknown emergence events have all begun with a wildlife or livestock spillover transmission event into humans, they each present unique subsequent public health challenges. Quantitative prediction of either the re-emergence of a known disease or potential for global spread of a novel disease can help optimize public health responses and resource allocation, but these events are usually analyzed in retrospect. In this dissertation, I developed quantitative frameworks that can be used in real-time for predicting disease emergence risk. In Chapter 2, I identified a seasonal trend to pandemic influenza emergence events, and proposed a hypothesis to explain the seasonal patterns and predict pandemic emergence risk for seasonal flu data. In Chapter 3, I developed a framework to both predict the number

of imported Zika cases into a region, and subsequently assist public health decision-making during an uncertain outbreak. Finally, in Chapter 4, I developed a method that can be used to update regional transmission risk estimates of a novel disease before transmission occurs. Altogether, the results presented in this dissertation suggest that statistical modeling can be an important tool to assist real-time public health predictions and responses.

Table of Contents

Chapter 4. Downgrading disease transmission risk estimates using terminal importations **75**

List of Tables

List of Figures

xix

Chapter 1: Introduction

Scientific and medical advances have greatly improved our ability to prevent and treat infectious diseases, but as of the year 2000, they still account for four of the top ten causes of mortality globally [187]. While malaria, Human Immunodeficiency Virus (HIV), and respiratory diseases are the largest continual threats globally, the World Health Organization (WHO) has prioritized diseases preparing for the next major new global epidemic, including a "Disease X" for currently unknown diseases [189]. Upon emergence in humans, public health officials will be forced to estimate the potential for a disease to spread locally and globally with little to no historical data to draw upon. For example, when the Zika virus emerged in the Americas, early estimates suggested that almost half of the United States might be at risk for transmission [1, 24]. However, even with over 5,000 imported cases, there have only been 229 locally acquired cases and these have been spatially limited to previously known hot spots in southern Texas and Florida (as of March 30, 2018). Had early predictions been more accurate, health officials could have properly allocated resources and the public could have planned travel accordingly.

Predicting novel emergences is inherently difficult, but statistical modeling can provide a quantitative framework to encapsulate potential sources of uncertainty and provide probabilistic forecasts to assist both public health pre-

paredness and response. In this dissertation I present three examples in which these methods can be used to predict disease emergence, quantify uncertainty, and provide insights for public health decision making.

In chapter 2, I attempt to better understand the risk for pandemic influenza emergence events. In looking back at historic pandemic emergence events, I uncovered an unexplained seasonal emergence trend: the six pandemics dating back to 1889 have all emerged in the spring or summer months. This was surprising, as influenza is normally thought of as a winter disease. I developed a mathematical model describing competition between seasonal and pandemic flu strains, and found close alignment between the model predictions and the actual historic pandemic emergence timing. Our model also made an interesting prediction concerning pandemic wave dynamics, which aligns well with many historical records. If our hypothesis holds, then public health officials will be better able to allocate surveillance resources, and anticipate and respond to emerging pandemics.

Chapter 3 lays a framework for how statistical models can directly assist public health decision-making in the face of an uncertain but growing epidemic. The framework was motivated by the emerging Zika epidemic in the Americas, where public health officials in the United States knew there would be many imported cases, but didn't know where and when to expect those cases. Furthermore, Zika had a notoriously low reporting rate [46, 50], meaning the detection of a single locally acquired case could indicate a single transmission event or a growing epidemic. The framework we developed pre-

dicted where and when importations would arrive in Texas in 2016, identified the Houston metropolitan region and the lower Rio Grande Valley as the most at risk regions for Zika in Texas, and provided a quantitative means for public health officials to develop county-specific *triggers* - guidelines for the number of detected cases needed before enacting public health interventions.

In carrying out the work in Chapter 3, I found that current methods for estimating transmission risk of vector-borne diseases are inadequate for public health decision-making. Predictions for county transmission risk suggested a wide-range of expected outbreak scenarios, from limited local transmission to full-blown epidemics. However, I discovered that there was information contained within the lack of transmission from the hundreds of importations that were found across Texas, and in Chapter 4 I developed a method to downgrade previous transmission risk estimates based on the presence or absence of transmission from imported cases. Our new transmission risk estimates suggest that no county is expected to be at risk for a Zika epidemic, and correctly predict the number of local transmission events detected in Texas in 2017. Our method will also work for future emerging diseases, as it allows one to update transmission risk estimates in real-time in the midst of an emerging epidemic.

Chapter 2: Seasonality in risk of pandemic influenza emergence

2.1 Abstract

Influenza pandemics can emerge unexpectedly and wreak global devastation. However, each of the six pandemics since 1889 emerged in the Northern Hemisphere just after the flu season, suggesting that pandemic timing may be predictable. Using a stochastic model fit to seasonal flu surveillance data from the United States, we find that seasonal flu leaves a transient wake of heterosubtypic immunity that impedes the emergence of novel flu viruses. This refractory period provides a simple explanation for not only the spring-summer timing of historical pandemics, but also early increases in pandemic severity and multiple waves of transmission. Thus, pandemic risk may be seasonal and predictable, with the accuracy of pre-pandemic and real-time risk assessments hinging on reliable seasonal influenza surveillance and precise estimates of the breadth and duration of heterosubtypic immunity.

2.2 Introduction

Influenza pandemics have emerged regularly throughout the 20th and 21st centuries, resulting in significant morbidity and mortality [92]. In preparation for future pandemics, public health agencies have enacted measures

4

to expedite pandemic vaccine development [193]. However, the manufacturing and distribution process is still expected to take several months, as occurred following the initial identification of the 2009 H1N1 pandemic virus [31, 43, 63, 99, 193]. In the interim, the primary pandemic control measures will include prophylaxis and treatment with antiviral medications and social distancing measures [81, 140, 193]. Given the potential severity of disease and rapid pace of emergence, advanced warning and early response are imperative. Thus, public health agencies have established extensive surveillance networks in humans, livestock, and wild bird populations [56, 153, 188]. While these systems are designed to identify potential pandemic threats as infections arise, researchers have also conducted mutatagenesis experiments to identify *upstream* evolutionary risks, that is, potential pathways toward human infectivity and virulence [76, 82]. However, the utility of such "gain-of-function" experiments has been disputed, particularly given the risks associated with handling highly virulent influenza viruses [109].

While public health agencies cannot yet anticipate the timing and location of the next pandemic, past pandemics may provide insight into spatiotemporal trends in risk. All recent pandemics emerged in the Northern Hemisphere in the spring and summer months (Fig 2.1): March (1918), April (1957, 2009), May (1889, 1977), and July (1968), though the 1977 pandemic virus was highly similar to a previously circulating virus, and thus thought to have emerged via accidental release from a laboratory [130, 152]. The 1889, 1977 and 1968 pandemics produced single epidemic waves, while the 1918, 1957, and 2009 pan-

5

demics spread in two waves–a relatively short spring-summer wave followed
by a more extensive fall wave [2, 3, 17, 18, 34, 37, 42, 87, 144, 168]. These pan-
demics also varied in severity, as measured by case fatality rates, with 1918
far more severe than the others [86, 172].
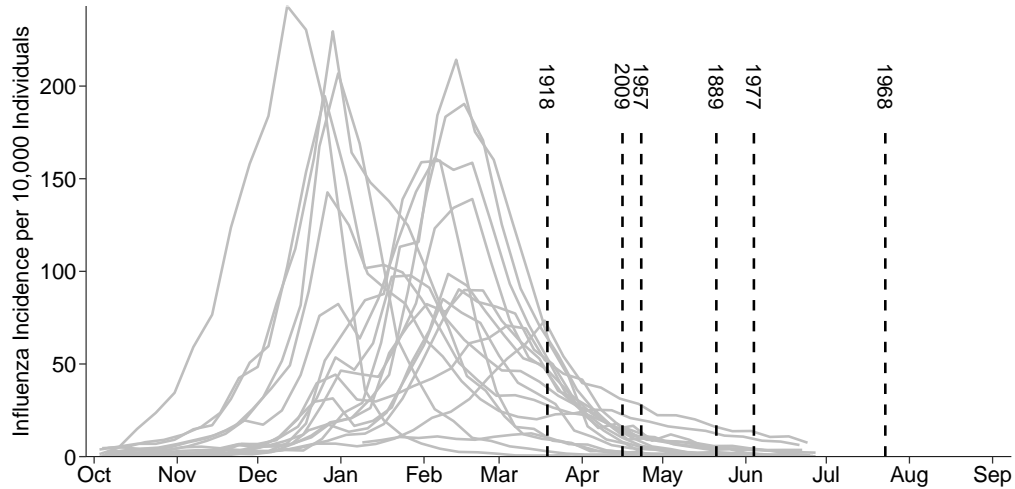
Figure 2.1: Historical pandemics emerged at the tail-end of flu seasons. Gray
curves show the 1995-2015 flu seasons in the US, excluding the 2009 H1N1 pan-
demic, as estimated by the CDC's ILINet surveillance system [190]. Vertical
dashed lines indicate emergence week of historical pandemics in their source
populations, defined as the first reported outbreak of severe influenza preced-
ing the initial pandemic wave. These estimates were obtained from: 1889
[144], 1918 [87, 168], 1957 [2, 3], 1968 [34, 42], and 1977 [17]. To be consistent,
we date the emergence of the 2009 pandemic according to the first significant
outbreak preceding the initial wave, rather than the earlier outbreaks in rural
Mexico that were identified only in retrospect [30].

The spring and summer emergence of the six recent pandemics seems
more than just a coincidence (Multinomial test; $p < 0.05$), but the sample
is quite small and derived from imperfect historical data. If, indeed, pan-

6

demic risk is seasonal, there are several plausible drivers. Two factors might favor pandemic emergence *during* the typical flu season. First, the socio-environmental conditions thought to promote seasonal influenza transmission (*e.g.*, humidity and school calendar) might also favor pandemic transmission during the winter months [116, 159, 160]. Second, pandemic emergence is often preceded by viral reassortment in hosts co-infected by a seasonal influenza virus and a novel virus, which should become more likely as the prevalence of seasonal flu increases [164, 165, 195]. On the other hand, transient cross-immunity from seasonal influenza infections may impede infection by novel viruses during the flu season. Together, these counterbalancing factors may produce a tight and predictable window for pandemic emergence.

Viruses of a common subtype (*e.g.* H3N2) are known to compete via *homosubtypic* immunity, producing stereotypical single branch influenza phylogenies [15, 16, 94, 95, 200]. However, the extent and mechanisms of competition among viruses of differing subtypes (e.g. a resident seasonal virus and a novel pandemic virus) via *heterosubtypic* immunity are not fully understood [57, 67, 68, 98, 106, 133, 157, 195]. A first childhood influenza infection may provide lifelong heterosubtypic immunity against subtypes within the same phylogenetic group (Group 1 includes H1, H2, and H5; Group 2 includes H3 and H7) [52, 67, 95, 104, 126, 145, 195]. In addition, any childhood or adulthood influenza infection may provide temporary, generalized heterosubtypic protection against other subtypes, lasting from a week to several months [57, 74, 89, 100, 101, 106]. This is perhaps mediated by cells surviving influenza

infection that exhibit an increased antiviral response and naturally turnover within a short period of time [74]. Immunity may not fully prevent infection, but individuals infected within this period experience reduced viral shedding, disease severity, and infection durations, likely reducing subsequent spread of the disease [100, 106, 133, 134].

Heterosubtypic immunity among influenza viruses would naturally lead to competition between subtypes, with the strength of the competition determined by the magnitude and duration of the immune response. Even if heterosubtypic immunity were short-lived, seasonal influenza may temporarily impede the emergence novel influenza subtypes. If a pandemic virus manages to emerge during this so-called refractory period, it would likely start slow and accelerate as residual immunity wanes.

Here, we characterize the impact of seasonal influenza on both the likelihood and magnitude of pandemic emergence events, mediated by transient heterosubtypic immunity following infection, and then integrate environmental constraints on flu transmission to estimate the seasonality of pandemic emergence risk. We fit two mathematical models to historical influenza data–one that assumes a homogeneous population and another that captures realistic heterogeneity in contact patterns–and simulate the introduction of novel influenza virus throughout the influenza season. We focus our analysis on the 2008-2009 seasonal epidemic, since it directly preceded the 2009 pandemic; for comparison, we also analyzed the larger 2003-2004 season (Figs 2.9, 2.10, and 2.11). As expected, the risk of pandemic emergence declines in the wake

of seasonal influenza, as does the effective reproduction number (early transmission rate) of any emerging pandemic. The seasonality of pandemic risk depends critically on the duration of immunity and the structure of the host population.

2.3 Methods

We developed a stochastic two-strain influenza transmission model that incorporates contact network structure, heterosubtypic immunity, and new estimates of the seasonal flu reproduction number to investigate the dynamics of pandemic emergence risk. We simulated thousands of novel pandemic virus introductions to estimate the changing probability of pandemic emergence and the R_{eff} upon emergence, as the flu seasons unfold.

2.3.1 Two-strain influenza transmission model

We included short-term heterosubtypic immunity using a two-strain SEIPR (Susceptible-Exposed-Infectious-Protected-Recovered) network model similar to [150] (Fig 2.6). All individuals are initially susceptible to both seasonal and pandemic influenza. Upon infection with one strain, individuals progress through the Exposed and Infectious classes; upon recovery, they enter a short period of complete protection from infection by the other strain, after which they regain full susceptibility to that strain. Our modeling framework does not allow simultaneous infection (co-infection) by both subtypes, as co-infection is thought to be relatively infrequent during concurrent epidemics

[167]. Close sequential infections can occur in the model, as some individuals transition through the protected class almost immediately (Fig 2.7). We modeled single influenza seasons, and thus assumed that recovered individuals are fully and permanently immune to their infecting strain, that there are no births or deaths, and that the network structure does not change over the course of a single simulation.

Infectious nodes infect susceptible neighbors at a per contact rate of β_i, where $i \in \{\text{seasonal}, \text{pandemic}\}$ indicates strain. We estimate β_seasonal by fitting a seasonal transmission model to recent flu season data (see section 2.3.3), and consider three transmission rate scenarios for the pandemic virus (1) equal transmissibility ($\beta_\text{pandemic} = \beta_\text{seasonal}$), (2) lower transmissibility ($\beta_\text{pandemic} < \beta_\text{seasonal}$), and (3) higher transmissibility ($\beta_\text{pandemic} > \beta_\text{seasonal}$).

We assume that durations of the exposed, infectious, and recovered periods are exponentially distributed. Upon infection by either strain, individuals instantaneously enter the Exposed class, then become infectious stochastically, at rate $\eta = \frac{1}{2.26}\text{days}^{-1}$, recover from infection at rate $\gamma = \frac{1}{3.38}\text{days}^{-1}$, and leave the heterosubtypic immune period at rate $\alpha = \frac{1}{42}\text{days}^{-1}$, based on published estimates [74, 180]. We considered a range of heterosubtypic immune periods (Fig 2.8), and herein report results based on a 42-day duration.

Adding Seasonal Forcing

We implemented seasonal forcing through a traditional humidity-forced influenza model estimating R_0 through time ($R_0(t)$) [197]:

$$R_0(t) = \exp(-180q(t) + \ln(R_{0\text{max}} - R_{0\text{min}})) + R_{0\text{min}}$$

Where $q(t)$ is the specific humidity at time t. We set $R_{0\text{min}} = 0.8$, as it is the lower bound estimate from [197], and solve for $R_{0\text{max}}$ through model fitting. We used equation 2.3 to convert between R_0 and β, obtaining $\beta(t)$ for model fitting and simulation purposes. We used the daily average specific humidity for the United States from 2000-2016 available from NOAA [136].

2.3.2 Simulation implementation

We simulated two-strain influenza epidemics using a stochastic Gillespie next-reaction algorithm built from EpiFire, a C++ network epidemic simulation library [64, 80]. We generated random contact networks with specified degree distributions using a configuration model algorithm [80]. For purposes of comparison, we assume that the homogeneous and empirical networks share the same mean degree of $\langle k \rangle$, with all nodes in the homogeneous network having degree exactly equal to $\langle k \rangle$ and the degrees in the empirical network randomly assigned according to an exponential distribution with rate $\frac{1}{\langle k \rangle}$. Based on published estimates for a large urban network, we assume $\langle k \rangle = 16$ [13].

For each *scenario*—combination of contact network, pandemic introduction time, β_{pandemic}, and immune period α—we ran 5,000 simulations. Each was seeded by infecting five randomly chosen individuals with the seasonal virus; at the designated introduction time, a single randomly chosen susceptible individual was infected by the pandemic virus. We terminated the simula-

tions once no individuals remained Exposed or Infectious. For each simulation we tracked the number of nodes in each class (Fig 2.6) and the *average excess degree in the susceptible portion of the network*, which is defined as follows. Consider only nodes currently susceptible to pandemic infection; call edges connecting such nodes *susceptible edges* and the number of such edges emanating from a susceptible node, the *susceptible degree* of that node. Imagine choosing a random susceptible edge and following it to one of its nodes. The average excess degree is the expected number of susceptible edges emanating from that susceptible node (other than the one along which we arrived), and is given by $\frac{\langle k^2_{\text{susceptible}} \rangle - \langle k_{\text{susceptible}} \rangle}{\langle k_{\text{susceptible}} \rangle}$, where $\langle k_{\text{susceptible}} \rangle$ and $\langle k^2_{\text{susceptible}} \rangle$ are the average susceptible degree and average squared susceptible degree in the network. Simulation source code can be accessed at https://github.com/sjfox/EpiFire.

2.3.3   Data and model fitting

To estimate the seasonal flu transmission rate, we fit a simple deterministic, network-based, ordinary differential equation (ODE) SEIR model of seasonal flu transmission to national influenza data from the United States. We chose 2008-2009 as it preceded the 2009 pandemic and 2003-2004 as a representative *large* season [190], and specifically analyzed weeks 1-15 of 2009 and week 45 of 2003 through week 3 of 2004. We estimated seasonal influenza incidence (denoted *ILI+*) throughout these periods by multiplying the CDC's ILINet estimates of influenza activity by WHO public health lab estimates of percent positive flu tests [190], as suggested by [66].

We implemented the Volz-Miller edge-based compartmental ODE model [26, 122, 185], which is given by following equations:

$$S(t) = \psi(\theta(t))$$

$$E(t) = 1 - S(t) - I(t) - R(t)$$

$$\dot{I}(t) = \eta E(t) - \gamma I(t)$$

$$\dot{R}(t) = \gamma I(t) \tag{2.1}$$

$$\dot{\phi_I} = \eta \left[\theta - \phi_S(0) \frac{\psi'(\theta)}{\psi'(1)} - \frac{\gamma(1-\theta)}{\beta} - \phi_R(0) \right]$$

$$- (\gamma + \beta + \eta)\phi_I$$

$$\dot{\theta} = -\beta \phi_I$$

The system can be understood by considering a test individual, u, which is a random individual in the network chosen at time, $t = 0$. θ is the overall probability that a given contact of u has not transmitted to u, and ϕ_S, ϕ_E, ϕ_I, and ϕ_R are the probabilities that the contact has not transmitted to u and is currently susceptible, exposed, infectious, or recovered, respectively. S, E, I, and R denote the proportion of the population in each state, and the parameters β, η, and γ correspond to the per contact rate of transmission, the rate of becoming infectious upon exposure, and the recovery rate, respectively. $P(k)$ describes the degree distribution and tells us the probability a random individual has degree k in the network. It follows that the average degree is given by $\langle k \rangle = \sum_k k P(k)$. $S(k, 0)$ is the probability a random individual of degree k is initially susceptible, which leads to the probability generating function describing the proportion of susceptible individuals in the population

13

(ψ) as, $\psi(x) = \sum_k S(k, 0) P(k) x^k$, where x is the probability a given contact of u has not transmitted to u.

We match the parameters in this model to our stochastic two-strain model, including the disease progression parameters (γ and η), network structures and the initial introduction of five infections in a population of 10,000 (that is, $I(0) = \phi_I(0) = \frac{5}{10000}$, $\phi_S(0) = 1 - \phi_I(0)$, $R = \phi_R(0) = \phi_E(0) = 0$, $\theta(0) = 1$, and $\phi_I = \phi_I(0)$). We solved the system of equations numerically using the *deSolve* package in R [146, 166].

To estimate both the per contact transmission rate, β_{seasonal}, and time of epidemic introduction for each network, we minimized the sum of the squared errors between the 2008-2009 *ILI+* data and the incidence predictions from the ODE model, using the *optim* function in R [146]. Given that different network structures lead to different rates of epidemic growth, the flexible epidemic start time allows tighter fitting of the models to the seasonal flu incidence data. For the 2008-2009 season, we estimated the epidemic start date to be November 15th and December 18th, 2008 for the homogeneous and empirical networks, respectively [146].

2.3.4 Analytic approximations of emergence probability and effective R_0

We derive mean field approximations of the emergence probabilities and effective R_0 using the process outlined in [121], which we outline briefly here. The generating function for the number of infected nodes in the first

14

generation of an outbreak is given by

$$f(x) = p_0 + p_1 x + \cdots + p_j x^j + \cdots$$

where p_j is the probability the index case infects j neighbors. More specifically,

$$p_j = \sum_{k=j}^{\infty} P(k) \int_0^1 \mathrm{Bi}(k, j, T) P(T) dT$$

with $\mathrm{Bi}(k, j, T)$ denoting the probability of j successful outcomes from k Bernoulli trials with probability of success equal to the transmissibility, T, defined by the probability distribution $P(T)$. (The probability distribution for T is defined by the randomly drawn recovery and infectious times in the Gillespie simulation.) A node of degree k that has just been infected has $k - 1$ possible neighbors to infect. The probability that this node infects j neighbors is given by

$$q_j = \frac{1}{\langle k \rangle} \sum_{k=j+1}^{\infty} k P(k) \int_0^1 \mathrm{Bi}(k - 1, j, T) P(T) dT$$

Similar to $f(x)$, $h(x) = \sum q_j x^j$ is the generating function for the number of infections caused by a non-index case, which leads to the equation

$$h(x) = \int_0^1 \frac{P(T)}{\langle k \rangle} \sum_{k=1}^{\infty} [1 + T(x - 1)]^{k-1} k P(k) dT$$

Ignoring finite size effects, the generating function for the number of infections g generations after the initial infection is $f(h^{g-1}(x))$ where $h^{g-1}(x)$ denotes composition of h with itself $g - 1$ times. The extinction probability is the probability that eventually there are 0 infections $\lim_{g \to \infty} f(h^{g-1}(0))$. Setting

$x_0 = \lim_{g\to\infty} h^{g-1}(0)$ we find that the emergence probability in a naïve network is given by

$$\mathcal{P} = 1 - f(x_0) \tag{2.2}$$

It also follows that the basic reproduction number (R_0) is given by

$$R_0 = \langle T \rangle \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \tag{2.3}$$

as originally shown in [119], where $\langle T \rangle$ is the average transmission probability. Under our continuous-time constant-rate assumptions, this is $\langle T \rangle = \frac{\beta}{\beta+\gamma}$.

2.3.5 Statistical analysis of simulated epidemics

For a given scenario, we restricted our analyses to simulations in which a seasonal epidemic actually occurred (defined as outbreaks with cumulative incidence reaching at least 5% of the population). We then estimated the pandemic emergence probability as the number of pandemic introductions that progressed into sustained outbreaks (infecting at least 5% of the population) divided by the number of simulations with seasonal epidemics of that scenario. We approximated the emergence timing of the pandemic as the first day in which the daily incidence was ≥ 5 individuals, as this was a good indicator for the beginning of the exponential growth phase.

Each time a pandemic successfully emerged, we estimated its R_{eff} by fitting the corresponding (empirical or homogeneous) single strain ODE network model (defined by equations in 2.1) to the simulated pandemic time series. The procedure is as described in the Data and model fitting section (2.3.3),

with two alterations: (1) we fix the introduction time to that specified by the simulation scenario and only estimate the transmission rate, and (2) we fit the model to the *cumulative incidence* of the pandemic virus. To then obtain an R_0 estimate, we plugged the estimated pandemic transmission rate and full degree distribution into equation 2.3.

During the *refractory period*, immunity in the population increases the transmissibility necessary for a pandemic to invade the population. We use equation 2.3 to estimate this changing invasion threshold; we let $\langle k^2 \rangle$ and $\langle k \rangle$ reflect the current susceptible portion of the network, set $R_0 = 1$, and solve for T. For a given time point t and scenario, we calculated T for a single, prototypical simulation and divided it by the comparable threshold in a completely susceptible population. Importantly, this analysis assumes that the network susceptibility is frozen in time at the introduction time, and does not take into account subsequent epidemic and pandemic dynamics.

2.4   Results

We fit two network models—an empirical model and a homogeneous model (roughly equivalent to a simple mass action model)—to influenza data from the 2008-2009 season in the US (Fig 2.2A), and estimated reproduction numbers (R_0) of 1.8 and 1.4, respectively (Analogous results for the larger 2003-2004 flu season are presented in section 2.6). Both estimates are consistent with prior studies [18, 186], and their discrepancy highlights a potential pitfall of simple epidemiological models. Given the observed heterogeneity in

human social behavior [13], the mass action models, which assume that all individuals have identical contact rates, may underestimate epidemic potential. Using these estimates of R_0, we simulate typical seasonal influenza epidemics and estimate the evolving probability of pandemic emergence. We assume an average 42-day period of complete heterosubtypic immunity upon recovery from a seasonal flu infection (Fig 2.7), which corresponds to the waning of generalized immunity in a human club cell-like line [74] and lies in between other estimates [57, 106]. We provide a sensitivity analysis with respect to the duration of immunity in the supplementary information (Fig 2.8).

2.4.1 Pandemic refractory period

Heterosubtypic immunity is expected to reduce pandemic emergence during the seasonal epidemic, with pandemic emergence probability reaching a minimum just after the epidemic peak of the flu season and then quickly rebounding (Fig 2.2B). The length and intensity of this *pandemic refractory period* should increase with the duration of heterosubtypic immunity, with prolonged immunity leading to complete pandemic exclusion (Fig 2.8).

The refractory period also depends on the transmissibility of the pandemic virus: the greater the transmission rate, the more readily a pandemic will emerge with or without immunological interference; the opposite is true for less transmissible viruses (Fig 2.14). The refractory effect is greater in the empirical (network) model than in the homogeneous model, suggesting that mass action assumptions may lead to underestimation of viral interference and
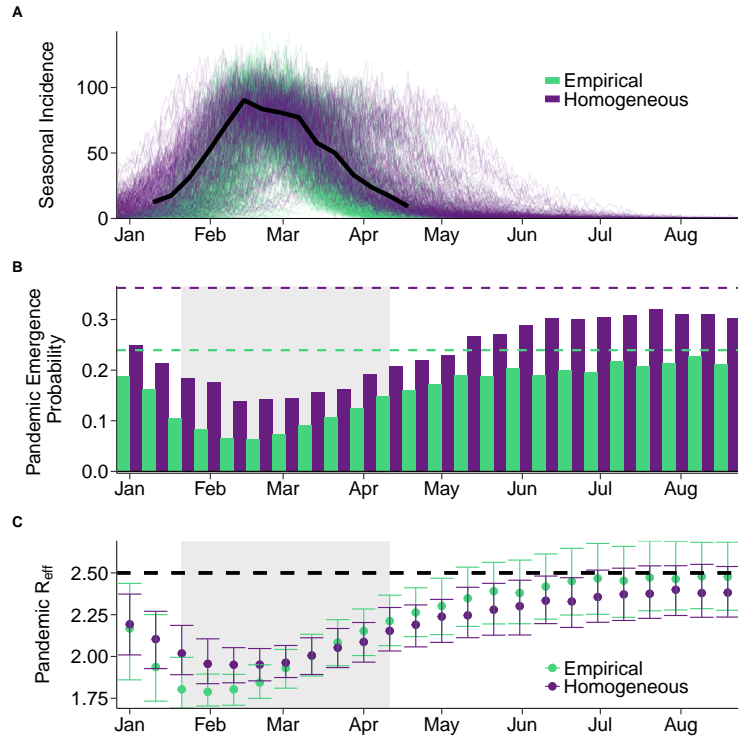
Figure 2.2: Seasonal epidemics produce a pandemic refractory period. A: Actual 2008-2009 epidemic curve (solid black line) and 200 stochastic simulations of seasonal epidemics for each network (green for empirical; purple for homogeneous), assuming transmission parameters estimated from 2008-2009 data. B: The probability of pandemic emergence upon the introduction of a single infected individual, assuming that the pandemic virus has the same transmission rate as the seasonal virus. Horizontal dashed lines indicate the emergence probabilities in a completely susceptible population calculated with Equation 2.2. The pandemic refractory periods (shaded regions) are expected to occur during and immediately following the seasonal epidemic peak. C: Assuming that the emerging pandemic has an $R_0 = 2.5$ in a naïve population (dashed horizontal line), we plot the median (points) and interquartile range of the measured R_{eff}, for each introduction time and network.

overestimation of pandemic risk. Assuming that the pandemic virus has the same intrinsic transmission rate as the seasonal virus, the probability of pandemic emergence is reduced by 73% and 62% in the empirical and homogeneous models, respectively, at the base of the refractory period, relative to comparable introductions in a completely susceptible population. Higher transmission rates lead to smaller reductions (56% and 19% respectively), while less transmissible viruses can be almost fully excluded (99% and 84% reductions respectively) (Fig 2.14). The 2008-2009 influenza season was relatively mild; larger seasonal flu epidemics produce deeper refractory periods, as illustrated for the 2003-2004 influenza season (Figs 2.9, 2.10, and 2.11).

2.4.2 Underestimation of pandemic R_0

For each simulated pandemic that successfully emerges, we estimate the effective R_0 (R_{eff}) of the virus, the reproduction number of the disease in a population that is not fully susceptible. Its magnitude depends on the extent of immunological interference by seasonal flu. Generally, the R_{eff} of the emerging pandemic virus decreases as the seasonal epidemic progresses towards its peak, bottoming out slightly before the emergence probability reaches its minimum. However, this occurs slightly earlier and more precipitously in the empirical model than in the homogeneous model (Fig 2.2C).

Whether or not a virus emerges depends on its intrinsic infectiousness and structure of the susceptible portion of the population. During the refractory period, the susceptible population is diminished, both in number and

20

connectivity. At the peak of the refractory period in the empirical network, we estimate that a introduced virus must be 1.16 times more infectious (transmissible) to emerge, relative to one entering a completely susceptible population. If the seasonal epidemics preceding the 1918 and 2009 pandemics were similar in timing and magnitude to our simulated epidemics, then we estimate that their intrinsic R_0's would have been 1.08-1.20 and 1.05-1.13 times larger, respectively, than their R_{eff}'s as the first waves emerged.

2.4.3 Contact networks determine pandemic vulnerability

The different levels of pandemic risk observed in our two models stem from their underlying contact networks. To illustrate this, we use *nodes* and *edges* to represent individuals and contacts between individuals, respectively. The *degree* of a node is defined as the number of edges connecting it to other nodes. The homogeneous model assumes that all individuals have the same number of contacts; the empirical model assumes realistic variation in degree [13]. We constrained the two models to have the same total number of nodes and empirically-derived mean degree, and, consequently, the same total number of edges.

The *susceptible* portion of a network is the subset of individuals that are currently susceptible to pandemic infection and any connections among them (Fig 2.3, orange circles and lines). (This is also known as the *residual* network [14].) The *susceptible degree* of a susceptible node is the number of contacts it has with other susceptibles. Upon infection by seasonal flu,

21

individuals and their coincident edges leave the susceptible network, returning only when their heterospecific influenza immunity wanes. This dynamic wake of immunity depends on the underlying network structure and, importantly, determines the population's vulnerability to pandemic emergence (Fig 3, grey nodes and edges).
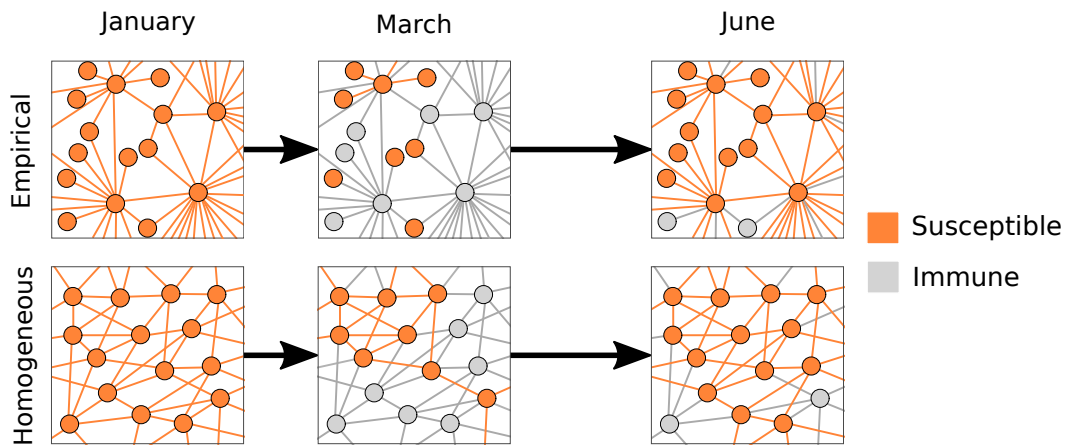
Figure 2.3: The evolving structure of the susceptible population as the flu season unfolds. For purposes of illustration, we present *caricatures* of each model through time, assuming that the average degree is $\langle k \rangle = 6$ and that we repeatedly observe the same subset of each population. Orange represents individuals susceptible to infection by the pandemic virus and the contacts between them; gray indicates individuals who are currently or recently infected by the seasonal virus, and thus immune to pandemic infection.

As a disease spreads, the chance of a node becoming infected will depend on its degree. The more contacts a node has, the higher its exposure risk. In a homogeneous network, chains of transmission progress randomly; in a heterogeneous network, outbreaks hit the most connected nodes earliest and hardest. Consider two emerging outbreaks–one in the empirical network

and another in the homogeneous network–that have reached the same cumulative incidence. Although the susceptible networks will have identical numbers of susceptible nodes, the empirical susceptible network will be much sparser (fewer edges) than the homogeneous counterpart, and will thus be more refractory to pandemic invasion (Fig 2.3, middle panels).

In a randomly selected pair of simulations, the homogeneous network decays to an susceptible network consisting of 71% of its original nodes and 43% of its original edges, before rebounding (Fig 2.4A). In contrast, the empirical network maintains more nodes (78%) and fewer edges (36%) at its most refractory moment, with the high degree nodes bearing the brunt of the seasonal epidemic (Fig 2.4B and C).

2.4.4   Seasonal pandemic emergence timing

The above analyses assume that pandemic emergence is constrained solely by heterosubtypic immunity, and do not consider the socio-environmental factors that shape seasonal flu dynamics. When we incorporate humidity-forced seasonality into the model, we find that pandemics are most likely to emerge soon after the seasonal epidemic peak (Fig 2.5), with the timing more constrained to the spring and early summer in the empirical network than in the homogeneous network. For both models, the most probable week of emergence falls within one week of the actual 2009 pandemic emergence event. To assess the consistency of our models with observed pandemics (five emerged in the spring and one emerged in the summer), we conduct multinomial tests of
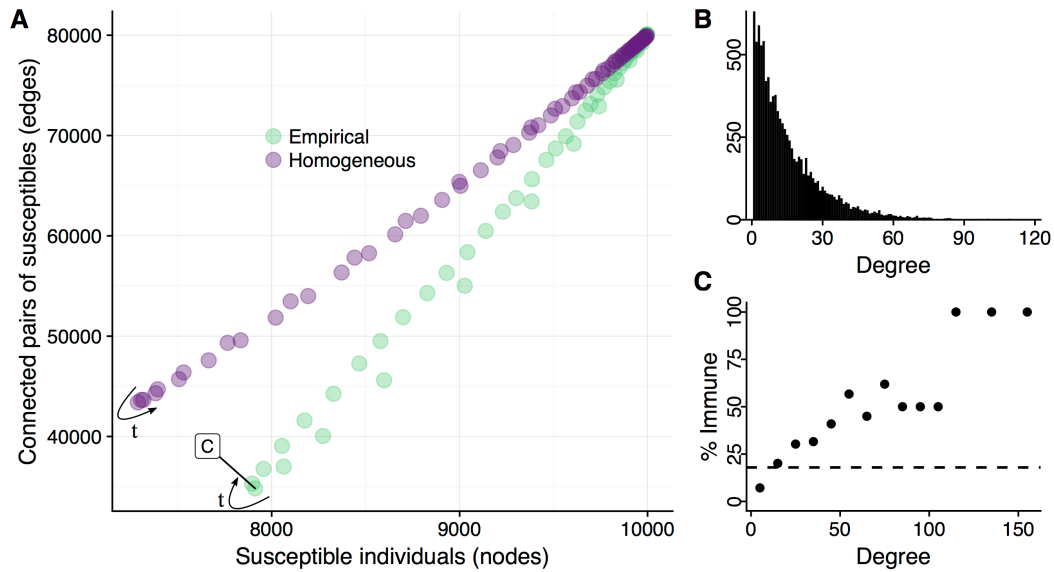
Figure 2.4: Seasonal flu disconnects the susceptible portion of a population. A: For a typical seasonal epidemic simulation, the number of individuals susceptible to infection by a pandemic virus and the number of edges connecting two such individuals are plotted for each network (green for empirical; purple for homogeneous), with each point representing a single time point over the course of the epidemic. Arrows indicate temporal progression. B: The distribution of degrees (number of contacts) assumed for the empirical model. The homogeneous model assumes that all individuals have 16 contacts. C: Snapshot of the susceptible portion of the empirical network at the base of the refractory period (at the time point indicated in panel A by the box labeled 'C'). Points indicate the percent of the nodes that are immune to pandemic infection, across different levels of connectivity. (We bin degrees by 10; for example, the lowest bin includes individuals with 1 to 10 contacts). For comparison, the horizontal dashed line indicates the overall proportion of individuals immunized in the network at the base of the refractory period.

the model-derived probabilities of emergence across each of the four seasons. While the empirical and homogeneous models are consistent with recent history (multinomial exact test $p = 0.53$ and $p = 0.35$, respectively), the simple

null model in which emergence risk is assumed to be constant throughout the year is not ($p < 0.05$). Although all historic pandemics seem consistent with the model, we note that these estimates are based on the 2008-2009 flu season and thus strictly pertain only to the 2009 pandemic. We speculate that projections from the seasons preceding each of the other historical pandemics would be similar and perhaps even better aligned with the emergence of the corresponding pandemic.

Earlier seasonal epidemics give rise to earlier risks of pandemic emergence (Fig 2.11), and extending the period of pandemic introduction from just the flu season to the entire year reduces the spring/summer emergence probability and renders the model predictions inconsistent with historic pandemic timing (Fig 2.12 ($p < 0.05$) and Fig 2.13 ($p < 0.05$)).

2.5 Discussion

The coincidental timing of recent pandemics may reflect multiple constraints, including environmental and behavioral factors that shape influenza's transmissibility (*e.g.* humidity, school calendar, etc.), reassortment events mediated by co-infection, and immune-mediated competition between pandemic and established viruses [57, 116, 156, 159, 160]. On the one hand, we would expect pandemics to emerge during the flu season, when socio-environmental conditions are conducive to influenza transmission and co-infections are likely; on the other hand, those would be the months of greatest competitive interference. These competing effects suggest that the risk of pandemic emergence

25
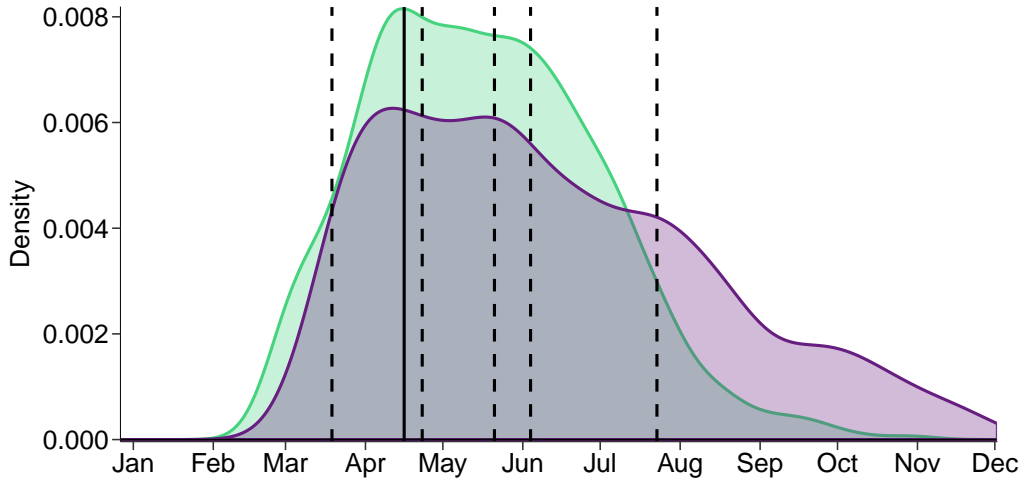
Figure 2.5: Seasonality further constrains pandemic emergence timing. Probability density for pandemic emergence timing for pandemics that emerge during the seasonal influenza epidemic for the homogeneous (purple) and empirical (green) networks. Pandemic emergence timing, the time in which the simulated pandemic begins rapid spread, is defined as the day the pandemic strain incidence reaches five or more cases. Results are for a pandemic emerging during the 2008-2009 flu season with the same transmission rate as the seasonal epidemic. Vertical lines indicate the timing of historic pandemics, with the solid line indicating the timing of the 2009 pandemic and dashed lines indicating timing of others.

may be greatest at the tail of the flu season, when conditions are still favorable and co-infections are possible, but competition is waning. Consider the following scenario. A novel virus is produced by co-infection mediated reassortment during the heart of the influenza season. It initially stutters, hindered by widespread heterosubtypic immunity, but does not completely disappear. With each new infection, emergence is possible and increasingly probable, as heterosubtypic immunity dissipates. This is consistent with the

26

timing of all recent pandemics (Fig 2.1), most of which (all but 1977) were caused by livestock-human reassortment viruses [155, 156, 165, 173, 195].

Pandemic emergence requires both the evolution of novel pandemic subtypes capable of human-to-human transmission and the ability of such new viruses to spread once they have appeared in humans. Our study focuses exclusively on the latter process, the success of new human-transmissible influenza viruses facing dynamic short-term heterosubtypic immunity resulting from seasonal influenza. Specifically, we have modeled a scenario in which potentially pandemic viruses appear (starting with a single infection) with constant probability during or following a typical Northern Hemisphere flu season. Individuals infected by seasonal flu are assumed to enjoy a short period of immunity towards other influenza subtypes, including the novel pandemic virus. Under reasonable assumptions regarding the duration of heterosubtypic immunity and human contact patterns, we characterize the changing risk of pandemic emergence throughout the flu season and find a considerable refractory period that is consistent with historical pandemic emergence events in the spring and summer months.

The rate of pandemic spread will depend on the time of emergence. Pandemics emerging during the seasonal refractory period will initially grow slowly and accelerate as residual immunity dissipates. Thus, the threat and pace of global expansion may far exceed projections based on early estimates of the viral reproduction number (R_0). In our model based on the 2008-2009 flu season, we estimate that, at the peak of the refractory period, naturally

27

occurring immunity will reduce the probability of pandemic emergence 73% and reduce the reproduction number of a successfully emerging virus by 30%. We assumed that all recovered individuals experience full protection during their short period of heterosubtypic immunity. If, for example, heterosubtypic immunity is incomplete or fails to prevent subsequent spread, the refractory effect may be diminished. Nonetheless, the qualitative results, including the timing of the refractory period and differences between the two network models should still hold.

Our comparison between homogeneous and empirical contact networks suggests that, while the refractory effect is robust, estimation of pandemic risk prior to and during emergence events will be highly sensitive to statistical assumptions regarding population structure. Several other studies have examined the impact of network structure on herd immunity following an epidemic or vaccination campaign, and similarly found that contact heterogeneity amplifies the refractory effect [14, 59, 103, 132]. Conventional models, that ignore social heterogeneity, are likely to overestimate both the emergence risk during the refractory period and the early transmission rate (R_0) of an emerging pandemic virus. Given the simplicity and growing flexibility of network methods, this further supports their scientific and public health utility [123, 124, 185].

Pandemics often emerge in multiple waves [107], including a herald wave in the spring or summer and a secondary wave the following fall or winter. These wave patterns are well documented for the 1918, 1957, and 2009 pandemics [11, 36–38, 48, 62, 131, 139]. Our results provide potential insight into

28

this phenomenon. The asynchronous forces of heterosubtypic herd immunity and suppressive off-season conditions may constrain pandemic emergence to the immediate wake of the flu season, exactly when lingering population-wide immunity is expected to dampen the initial wave of pandemic transmission. In the months following, the limited herald wave runs its course, residual seasonal immunity continues to decline, and socio-environmental conditions slowly become more conducive to flu transmission, thereby setting the stage for a major winter pandemic wave. Early estimates of pandemic R_0 that do not properly account for underlying population immunity may substantially underestimate the magnitude of the second (fall or winter) pandemic wave, as the R_{eff} at the time of emergence may be considerably lower than R_0 in a fully susceptible population. Our analysis suggests that a pandemic emerging between March and June may produce a secondary wave with an R_{eff} that is 4-28% larger than the initial R_{eff}, depending on the duration of heterosubtypic immunity, the timing of emergence, and the baseline transmissibility of the virus. Recent analyses of the 1918 and 2009 pandemic waves found that the initial waves were 3.6% and 6.5% less transmissible than the secondary wave, supporting our conclusions [197, 198]. This finding is broadly consistent with published estimates for the reproduction numbers of primary and secondary pandemic waves, with the exception of the 1918 pandemic in Denmark [11, 18]. This discrepancy may be attributable to poor data or stem from local differences in the preceding flu season or population structure.

Most historic pandemic viruses were likely created by recent livestock-

human reassortment events [164, 165, 195], with two possible exceptions. The 1977 pandemic was caused by a lab escapee, and the 2009 pandemic evolved from a human-derived variant that circulated in swine for years before the precipitating reassortment event, which may have occurred in livestock several months or years prior to its 2009 emergence [165]. We have assumed that pandemic introductions will be constrained to the flu season for two reasons (Fig 2.5). First, the chance of a livestock-human reassortment event will depend on the prevalence of flu in both humans and livestock, and thus increase as seasonal flu gains momentum. Second, flu prevalence in livestock is thought to mirror seasonal flu in humans [45, 115]. Thus, even viruses emerging directly from livestock, without a precipitating human reassortment event, may be constrained to the same months.

When we remove this assumption and introduce pandemic viruses throughout the year, the plausible emergence times start earlier in the fall, before the seasonal flu epidemic takes hold (Fig 2.12 and Fig 2.13). This broader emergence scenario is inconsistent with historical pandemics, given that none emerged in the fall just prior to a seasonal epidemic. While this does not prove our seasonality assumptions, it suggests that there may be factors restricting emergence events in the fall, such as the ones we have hypothesized, or that fall emergence simply has not occurred, by chance alone, across the limited number of recent pandemics. Our study was not designed to detect such seasonal constraints on pandemic emergence (rather we assume them and analyze the consequences), but leaves this as an interesting open question for future

work. Interestingly, the broader emergence scenario may apply to the 1977 pandemic, which, unlike the other recent pandemics, did not emerge directly from influenza circulating in livestocks or humans, as well as to risks associated with future gain-of-function avian influenza experiments.

This approach can be readily applied to other retrospective or predictive global risk assessments, using seasonal flu surveillance data at the relevant geographic and temporal scale [90, 170]. Our results suggest that Southern hemisphere pandemic risk will be greatest in September and October following their respective flu season [60]. Tropical and subtropical regions, which have low levels of sporadic flu transmission, seasonal patterns, or bimodal seasonality should experience refractory periods in the wakes of their respective epidemics [10, 161, 170, 171, 184]. Estimating spatiotemporal emergence risks will require data-driven models that consider local flu seasonality and contact networks, both of which can vary greatly with climatic zone and human developmental index. Such analyses can support pandemic planning, including the targeting of surveillance systems for detecting emergence events around the globe [108, 153].

Our model makes several assumptions about the transmissibility of both seasonal and pandemic influenza viruses. We assume that the intrinsic transmission rates depend on humidity (Fig 2.5), and do not explicitly consider other environmental and sociological factors that may be important (e.g., school calendar) [83, 116]. Since we estimated pandemic emergence locations and dates based on reports of major outbreaks, our estimates may be

31

biased towards regions with high reporting rates or population densities. Our study is further limited by the small number of pandemic emergence event; with five natural pandemics emerging in subtropical and temperate climates, we lack the power to fit high resolution predictive model. Instead, we used the North American 2008-2009 influenza season as a prototypical flu season for exploring seasonal and immunological drivers of pandemic risk. The flu seasons preceding the other 20th century pandemics likely varied in both timing and magnitude. Additional historical data from those pandemics and their preceding seasons might enable more reliable spatiotemporal estimates of pandemic emergence risk and the duration of cross-immunity.

Recent pandemics exhibited similar timing and geographic origins, having all emerged in the Northern Hemisphere. Why this is so, and whether it suggests higher risk of future pandemic emergence in the Northern Hemisphere is yet unknown. Molecular analyses suggest that seasonal flu diversity is seeded in the Northern Hemisphere (Southeast Asia) [16]. Furthermore, human and livestock populations tend to have higher densities in the Northern Hemisphere than the Southern Hemisphere [111, 149]. These two factors could suggest that the Northern Hemisphere may be a likely source for future pandemics. If influenza refractory periods are estimated for other climatic zones, as we have done here for the Northern Hemisphere, we may better understand the common origins of past pandemics and gain actionable insights into global dynamics of pandemic risk.

We also focus exclusively on transient heterosubtypic immunity imme-

diately following seasonal flu infection, which is only one of many forms of immunological heterogeneity that may constrain pandemic emergence. For example, the age-specific rate of severe and deadly infections of novel H7N9 and H5N1 in China reflect long-lasting heterosubtypic immunity stemming from early childhood influenza infections [67]. Our model does not consider such long-term heterogeneity in susceptibility, nor does it consider intrinsic heterogeneity in heterosubtypic immunity following infection (e.g., variation in severity, transmissibility, or infectious period) [133, 134]. Incorporating historically-acquired immunity, individual heterogeneity, and future advances in our understanding of transient heterosubtypic immunity should improve pandemic risk assessments.

Our study is intended as a proof of concept. Using simple, conservative models of influenza transmission, cross-immunity, and seasonality, we lend support to a parsimonious explanation for the historical spring-summer timing of pandemic emergence and demonstrate that pandemic risk may be both seasonal and predictable. However, there is much we still cannot predict, such as when and where reassorted viruses capable of human-to-human transmission will arise. Recent human outbreaks of H7N9 and H5N1 influenza during the winter and spring months suggest that other factors may inhibit spread, such as intrinsic transmissibility and the underlying immunological landscape [51, 67, 97]. Although we do not address the biogeographic risks of novel viruses first arising through reassortment events in humans or livestock, laboratory experimentation, or other mechanisms, our study provides insight into the

subsequent risk of emergence and a method for estimating such risk from seasonal flu surveillance data. As we gain a better understanding of breadth and duration of heterosubtypic immunity, both in general and between specific combinations of influenza viruses, our insights and methodology can be applied to improve global surveillance, detection, planning and intervention efforts for pandemic influenza.

2.6 Supplemental Information

Parameters

β_1	Disease 1 per edge transmission rate
β_2	Disease 2 per edge transmission rate
η	1/incubation period = 1/3.38 days
γ	1/infectious period = 1/2.26 days
α	1/protected period = 1/42 days

Individual States

S	Fully Susceptible	S_{12}	Susceptible d2 post d1
E_{01}	Exposed disease 1	S_{21}	Susceptible d1 post d2
E_{02}	Exposed disease 2	E_{12}	Exposed d2 post d1
I_{01}	Infectious disease 1	E_{21}	Exposed d1 post d2
I_{02}	Infectious disease 2	I_{12}	Infectious d2 post d1
P_1	Protected	I_{21}	Infectious d1 post d2
P_2	Protected	R	Fully Recovered

$N = 10{,}000 = S + E_{01} + E_{02} + I_{01} + I_{02} + P_1 + P_2 + S_{12} + S_{21} + E_{12} + E_{21} + I_{12} + I_{21} + R$

$(n_{I_{01}} + n_{I_{21}})\beta_1$

$(n_{I_{02}} + n_{I_{12}})\beta_2$

$S \to E_{01} \xrightarrow{\eta} I_{01} \xrightarrow{\gamma} P_1 \xrightarrow{\alpha} S_{12} \to E_{12} \xrightarrow{\eta} I_{12} \xrightarrow{\gamma} R$

$S \to E_{02} \xrightarrow{\eta} I_{02} \xrightarrow{\gamma} P_2 \xrightarrow{\alpha} S_{21} \to E_{21} \xrightarrow{\eta} I_{21} \xrightarrow{\gamma} R$

$(n_{I_{02}} + n_{I_{12}})\beta_2$

$(n_{I_{01}} + n_{I_{21}})\beta_1$

Figure 2.6: Short-term heterosubtypic immunity model description for a single individual (node) in the network. Solid arrows indicate the individual's transitions through epidemiological states, and dashed arrows indicate neighbor influence on the individual's transitions, with $n_{I_{XY}}$ indicating the number of the individual's neighbors who are currently in state I_{XY}. Symbols labeling arrows indicate the transition rates between states (solid arrows), or the rate at which individuals transmit to susceptible individuals (dashed arrows). For example, an individual in state S_{21} has been infected and recovered from disease 2 and is currently susceptible to disease 1, so this individual will become exposed to disease 1 at rate $(n_{I_{01}} + n_{I_{21}})\beta_1$, where β_1 is the per contact rate of transmission for disease 1, and $n_{I_{01}} + n_{I_{21}}$ is the number of its neighbors who are currently infected with disease 1.
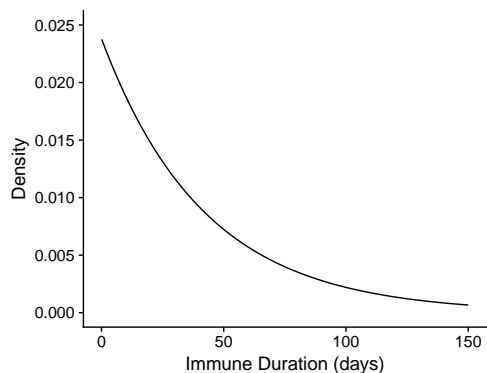
Figure 2.7: Probability distribution for immune duration. We model the immune duration as an exponentially distributed random variable with rate = 1/42, meaning the most likely immune duration is nearly zero days of immunity, but on average individuals will spend 42 days in the immune state. As the influenza epidemics we model last 100 days or more, the immune duration allows for an individual to experience serial infections of the seasonal and pandemic strain.
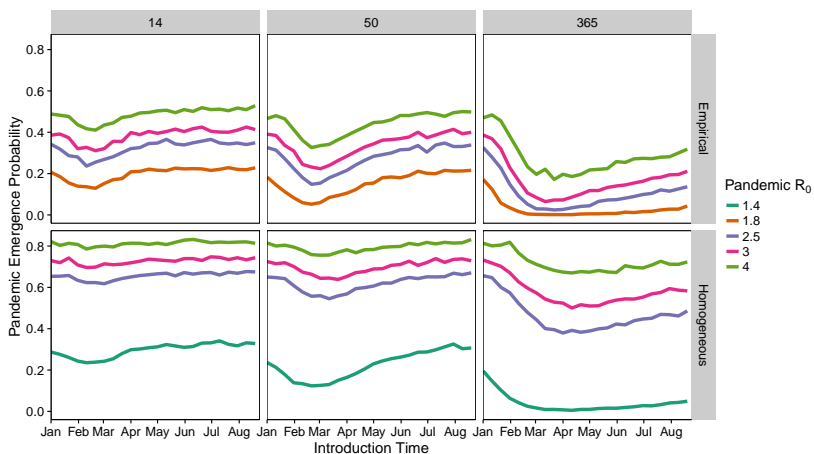
Figure 2.8: The extent and magnitude of the pandemic refractory period depends on the duration of cross-immunity. Columns represent the duration of cross-immunity (α), and the rows represent the two networks considered. Lines represent the emergence probability of pandemics across the 2008-2009 seasonal influenza epidemic for a variety of pandemic $R_0 s$ (colors).
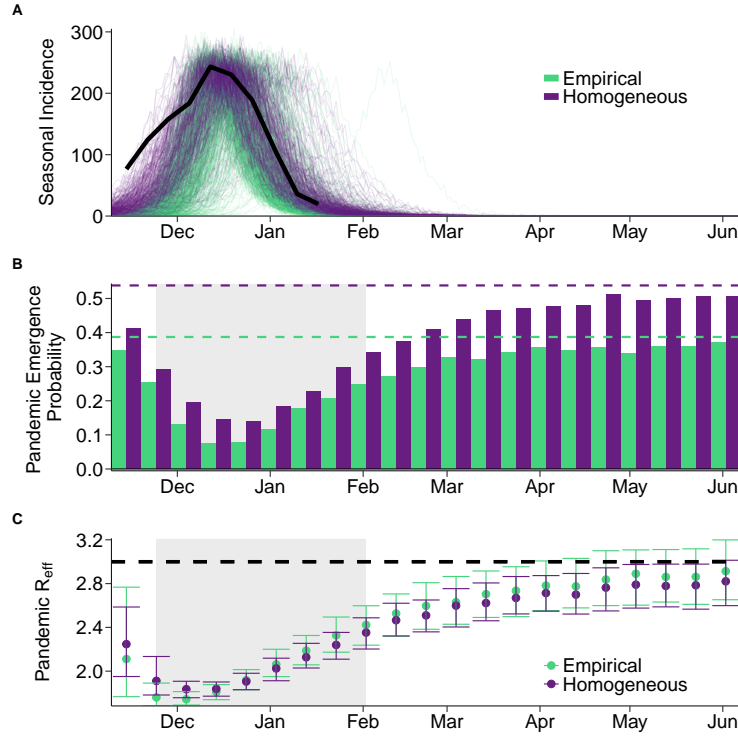
Figure 2.9: Larger seasonal epidemics produce larger pandemic refractory periods. A: Actual 2003-2004 epidemic curve (solid black line) and 200 stochastic simulations of seasonal epidemics for each network (green for empirical; purple for homogeneous), assuming transmission parameters estimated from 2003-2004 data. B: The probability of pandemic emergence upon the introduction of a single infected individual, assuming that the pandemic virus has the same transmission rate as the seasonal virus. Horizontal dashed lines indicate the emergence probabilities in a completely susceptible population calculated with Equation 2.2. C: Underestimation of pandemic R_0. Assuming that the emerging pandemic has an $R_0 = 3$ in a naïve population (dashed horizontal line), we plot the median (points) and interquartile range of the measured R_{eff}, for each introduction time and each network.
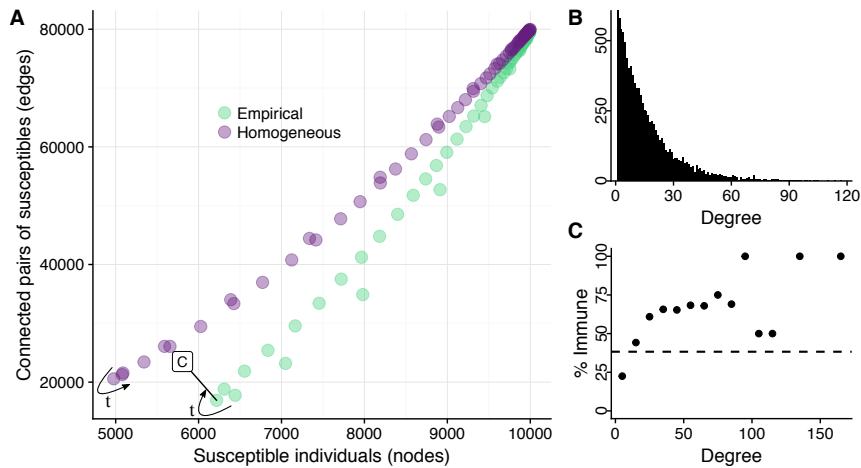
Figure 2.10: Seasonal flu disconnects the susceptible portion of a population (Large seasonal epidemic). A: For a single (typical) 2003-2004 seasonal epidemic simulation, the number of individuals susceptible to infection by a pandemic virus and the number of edges connecting two such individuals are plotted for each network (green for empirical; purple for homogeneous), with each point representing a single time point over the course of the epidemic. Arrows indicate temporal progression. B: The distribution of degrees assumed for the empirical model. The homogeneous model assumes that all individuals have 16 contacts. C Snapshot of the susceptible portion of the empirical network at the base of the refractory period (at the time point indicated in panel A by the box labeled 'C'). Points indicate the percent of the nodes that are immune to pandemic infection, across different levels of connectivity.

38
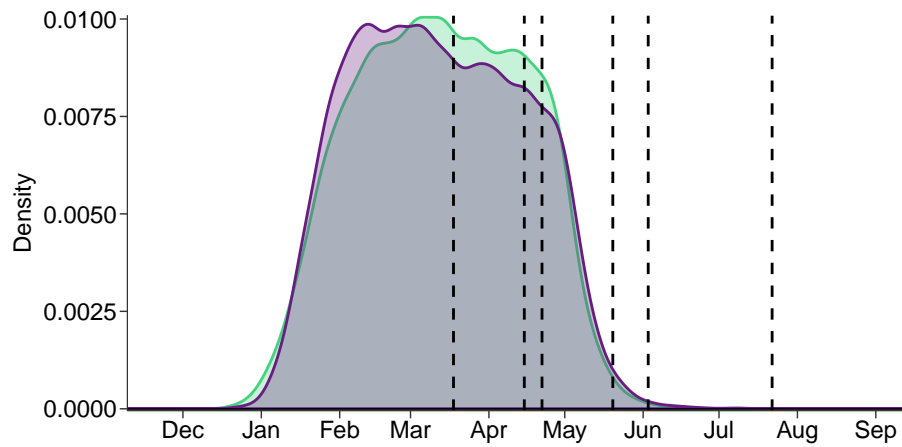
Figure 2.11: Seasonality also constrains pandemic emergence timing for the 2003-2004 season. Probability density for pandemic emergence timing for pandemics that emerge during the seasonal influenza epidemic for the homogeneous (purple) and empirical (green) networks. Pandemic emergence timing, the time in which the simulated pandemic begins rapid spread, is defined as the day the pandemic strain incidence reaches five or more cases. Results are for a pandemic emerging during the 2003-2004 flu season with the same transmission rate as the seasonal epidemic. Vertical dashed lines indicate the timing of historic pandemics.
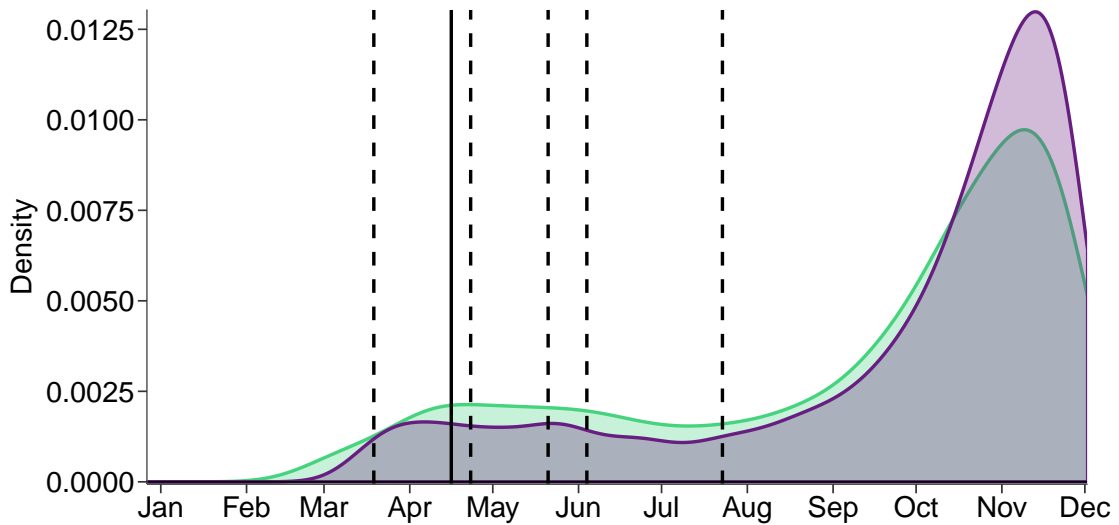
Figure 2.12: Pandemic emergence less constrained if pandemics introduced throughout the year (2008-2009). Probability density for pandemic emergence timing for pandemics that emerge across the whole year during and following the seasonal epidemic for the homogeneous (purple) and empirical (green) networks. Pandemic emergence timing, the time in which the simulated pandemic begins rapid spread, is defined as the day the pandemic strain incidence reaches five or more cases. Results are for the 2008-2009 flu season with the same transmission rate as the seasonal epidemic. Vertical lines indicate the timing of historic pandemics, with the solid line indicating the timing of the 2009 pandemic and dashed lines indicating timing of others.
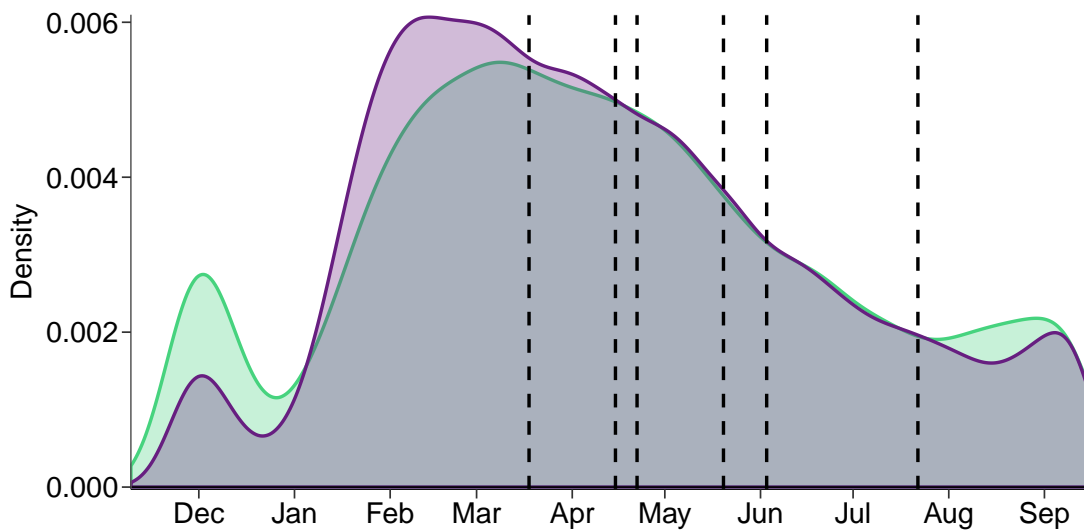
Figure 2.13: Pandemic emergence less constrained if pandemics introduced throughout the year (2003-2004). Probability density for pandemic emergence timing for pandemics that emerge across the whole year during and following the seasonal epidemic for the homogeneous (purple) and empirical (green) networks. Pandemic emergence timing, the time in which the simulated pandemic begins rapid spread, is defined as the day the pandemic strain incidence reaches five or more cases. Results are for the 2003-2004 flu season with the same transmission rate as the seasonal epidemic. Vertical dashed lines indicate the timing of historic pandemics.
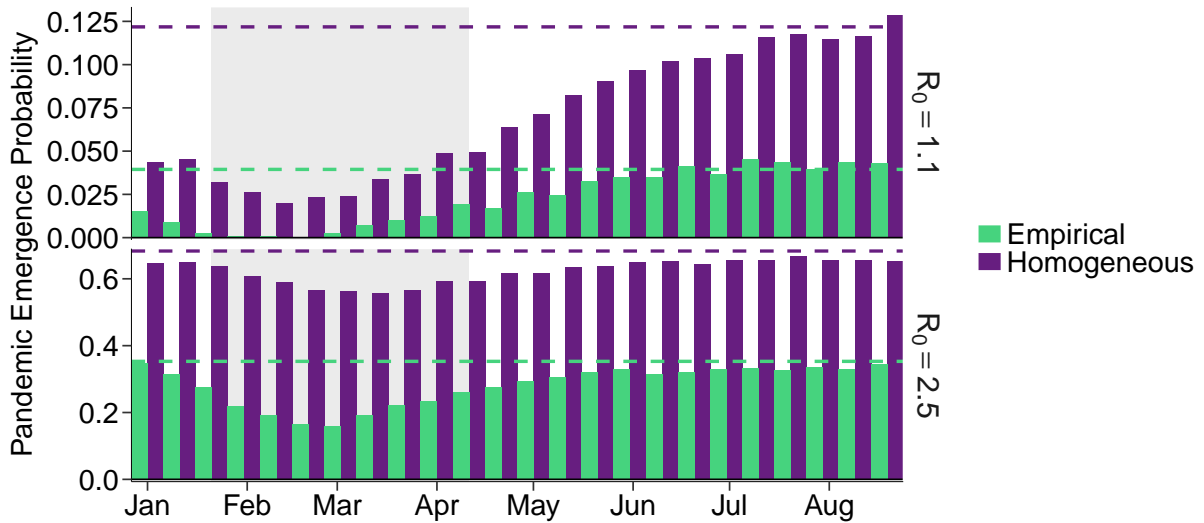
Figure 2.14: Pandemic refractory period reduces as transmissibility increases. Pandemic emergence probabilities plotted for the 2008-2009 seasonal simulation for a pandemic that is less transmissible than the seasonal strain (Top) and one that is more transmissible than the seasonal strain (Bottom) on the two analyzed networks (fill colors). Probability is estimated as the proportion of introductions that subsequently infected at least 5% of the overall population out of the 5,000 simulations. Horizontal dashed lines indicate the emergence probabilities in a completely susceptible population calculated with Equation 2. The pandemic refractory periods (shaded regions) are plotted the same as in the manuscript. Refractory period is deeper and wider for the less transmissible strain, and nearly disappears if the pandemic transmissibility is high enough.

Chapter 3: Assessing real-time Zika risk in the United States

3.1 Abstract

Confirmed local transmission of Zika Virus (ZIKV) in Texas and Florida have heightened the need for early and accurate indicators of self-sustaining transmission in high risk areas across the southern United States. Given ZIKVs low reporting rates and the geographic variability in suitable conditions, a cluster of reported cases may reflect diverse scenarios, ranging from independent introductions to a self-sustaining local epidemic. We present a quantitative framework for real-time ZIKV risk assessment that captures uncertainty in case reporting, importations, and vector-human transmission dynamics. We assessed county-level risk throughout Texas, as of summer 2016, and found that importation risk was concentrated in large metropolitan regions, while sustained ZIKV transmission risk is concentrated in the southeastern counties including the Houston metropolitan region and the Texas-Mexico border (where the sole autochthonous cases have occurred in 2016). We found that counties most likely to detect cases are not necessarily the most likely to experience epidemics, and used our framework to identify triggers to signal the start of an epidemic based on a policymakers propensity for risk. This framework can inform the strategic timing and spatial allocation of public health

resources to combat ZIKV throughout the US, and highlights the need to develop methods to obtain reliable estimates of key epidemiological parameters.

3.2 Introduction

In February 2016, the World Health Organization (WHO) declared Zika virus (ZIKV) a Public Health Emergency of International Concern [71]. Though the Public Health Emergency has been lifted, ZIKV still poses a great threat for reemergence in susceptible regions in seasons to come [?]. In the US, the 268 reported mosquito-borne autochthonous (local) ZIKV cases occurred in Southern Florida and Texas, with the potential range of a primary ZIKV vector, Aedes aegypti, including over 30 states[1, 61, 177]. Of the 2,487 identified imported ZIKV cases in the US through the end of August, 137 had occurred in Texas. Given historic small, autochthonous outbreaks (ranging from 4 - 25 confirmed cases) of another arbovirus vectored by Ae. Aegypti-dengue (DENV) [175–177], Texas was known to be at risk for autochthonous arbovirus transmission, and the recent outbreaks have highlighted the need for increased surveillance and optimized resource allocation in the states and the rest of the vulnerable regions of the Southern United States.

As additional ZIKV waves are possible in summer 2017, public health professionals will continue to face considerable uncertainty in gauging the severity, geographic range of local outbreaks, and appropriate timing of interventions, given the large fraction of undetected ZIKV cases (asymptomatic) and economic tradeoffs of disease prevention and response [8, 102, ?, 113]. De-

pending on the ZIKV symptomatic fraction, reliability and rapidity of diagnostics, importation rate, and transmission rate, the detection of five autochthonous cases in a Texas county, for example, may indicate a small chain of cases from a single importation, a self-limiting outbreak, or a large, hidden epidemic underway (Fig 3.1). These diverging possibilities have precedents. In French Polynesia, a handful of ZIKV cases were reported by October 2013; two months later an estimated 14,000-29,000 individuals had been infected [102, 113]. By contrast, Anguilla had 17 confirmed cases from late 2015 into 2016 without a subsequent epidemic, despite large ZIKV epidemics in surrounding countries [142]. To address the uncertainty, the CDC issued guidelines for state and local agencies; they recommend initiation of public health responses following local reporting of two non-familial autochthonous ZIKV cases [32].

Previous risk assessments of ZIKV have provided static *a priori* assessments based on historical incidence and vector suitability, but they do not provide dynamic risk assessments as cases accumulate in a region. Here, we present a framework to support real-time risk assessment, and demonstrate its application in Texas. Our framework accounts for the uncertainty regarding ZIKV epidemiology, including importation rates, reporting rates, local vector populations, and socioeconomic conditions, and can be readily updated as our understanding of ZIKV evolves. To estimate current and future epidemic risk from real-time ZIKV case reports, the model incorporates a previously published method for estimating local ZIKV transmission risk and a new model
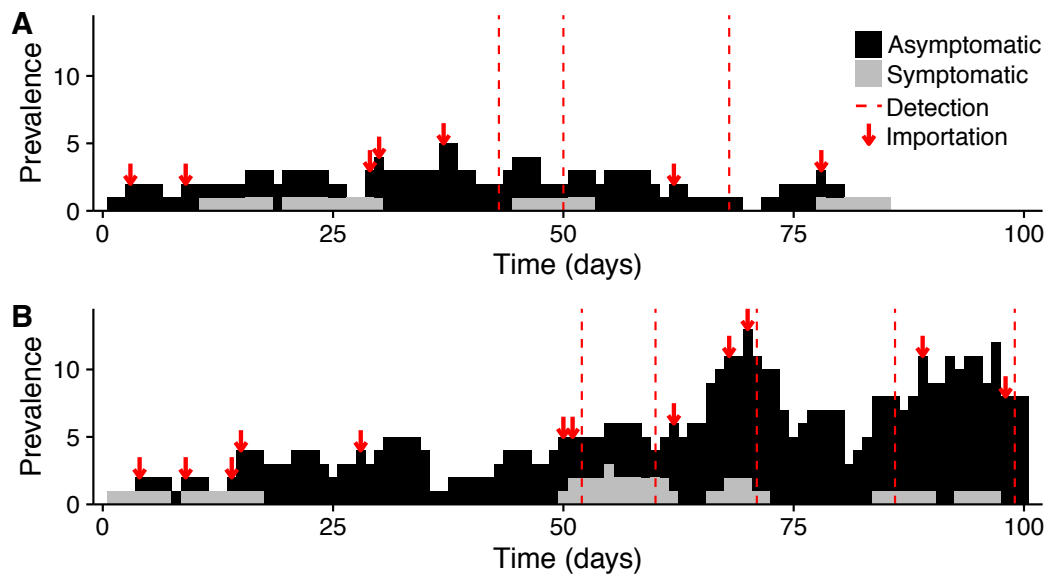
Figure 3.1: ZIKV emergence scenarios. A ZIKV infection could spark (A) a self-limiting outbreak or (B) a growing epidemic. Cases are partitioned into symptomatic (grey) and asymptomatic (black). Arrows indicate new ZIKV importations by infected travelers and vertical dashed lines indicate case reporting 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.

for estimating local importation risk. Across Texas 254 counties, we find that the estimated risk of a locally sustained ZIKV outbreak rises precipitously as autochthonous cases accumulate, and that counties at the southern tip of the Texas-Mexico border and in the Houston Metropolitan Area are at the highest risk for ZIKV transmission. This statewide variation in risk stems primarily from mosquito suitability and socio-environmental constraints on ZIKV

transmission rather than heterogeneity in importation rates.

3.3 Methods

Our risk-assessment framework is divided into three sections: (1) county-level epidemiological estimates of ZIKV importation and relative transmission rates, (2) county-specific ZIKV outbreak simulations, and (3) ZIKV risk analysis (Fig 3.5). To demonstrate this approach, we estimate county-level ZIKV risks throughout the state of Texas for August 2016, given that, by May 2016, Texas experienced dozens of ZIKV importations without subsequent vector-borne transmission.

3.3.1 Estimating Importation Rates

Our analysis assumes that any ZIKV outbreaks in Texas originate with infected travelers returning from active ZIKV regions. To estimate the ZIKV importation rate for specific counties, we (1) estimated the Texas statewide importation rate (expected number of imported cases per day) for August 2016, (2) estimated the probability (import risk) 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 ZIKV travel-associated cases were reported in Texas [177], yielding a baseline first quarter estimate of 0.3 imported cases/day throughout Texas. In 2014 and 2015, arbovirus introductions into Texas increased threefold over this same time period,

perhaps driven by seasonal increases in arbovirus activity in endemic regions and the approximately 40% increase from quarter 1 to quarter 3 in international travelers to the US [138]. Taking this as a baseline (lower bound) scenario, we projected a corresponding increase in ZIKV importations to 0.9 cases/day (statewide) for the third quarter.

2. To build a predictive model for import risk, we fit a probabilistic model (maximum entropy) [84] of importation risk to 183 DENV, 38 CHIKV, and 31 ZIKV Texas county-level reported importations from 2002 to 2016 and 10 informative socioeconomic, environmental, and travel variables. 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 [129]. 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 less than 10 reported ZIKV importations along the border to date (October 2016). We included DENV and CHIKV importation data in the model fitting so as to consider potential future importation pressure from Mexico, as ZIKV

continues its increasing trend since March 2016 [143]. 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 of training and testing sets of data [117, 192], reducing the original set of 72 variables to 10. We validated our importation model by comparing the predicted distribution of cases across the state given a total number of imported cases (September 2016) as a linear predictor of the empirical distribution of cases across counties.

3.3.2 County Transmission Rates (R_0)

The risk of ZIKV emergence following an imported case will depend on the likelihood of mosquito-borne transmission. For emerging diseases like ZIKV, the public health and research communities initially face considerable uncertainty in the drivers and rates of transmission, given the lack of field and experimental studies and epidemiological data, and often derive insights through analogy to similar diseases. For our case study, we estimated county-level ZIKV transmission potential by Ae. aegypti using a recently published model [7], that derives some of its key parameters from DENV data 3.1. The utility of our framework depends on the validity of such estimates and will increase as our knowledge of ZIKV improves. However, we expect our results to be robust to most sources of uncertainty regarding ZIKV and DENV epidemiology, as they may influence the absolute but not relative county-level

49

risks.

We estimated the ZIKV reproduction number (R_0), the average number of secondary infections caused by a single infectious individual in a fully susceptible population, for each Texas county following the method described in Perkins et al [7]. The method calculates R_0 using a temperature-dependent formulation of the Ross-Macdonald model, where mosquito mortality rate (μ) and extrinsic incubation period of ZIKV (n) are temperature dependent functions; the human-mosquito transmission probability ($b = 0.4$), number of days of human infectiousness ($\frac{c}{r} = 3.5$), and the mosquito biting rate ($a = 0.67$) are held constant at previously calculated values [4, 7, 19, 22, 33, 128]; and the economic-modulated mosquito-human contact scaling factor (m) is a function of county mosquito abundance and GDP data fit to historic ZIKV seroprevalence data [7]. To account for uncertainty in the temperature-dependent functions (the extrinsic incubation period (EIP) and mosquito mortality rate) and in the relationship between economic index and the mosquito-to-human contact rate, Perkins et al. generated functional distributions via 1000 Monte Carlo samples from the underlying parameter distributions. We assume DENV estimates for these temperature-dependent functions, since we lack such data for ZIKV and these Flaviviruses are likely to exhibit similar relationships between temperature and EIP in Ae. Aegypti [19]. We used the resulting distributions to estimate R_0 for each county, based on county estimates for the average August temperature, mosquito abundance from Kraemer et al [93], and GDP [19]. Our R_0 estimates were similar to those reported by Perkins et al. [7]

Parameter	Description	Values Investigated (or median 95%)	Source
Exposed compartments (e)	Number of exposed compartments	6	Fit (See Section 3.6.1)
Incubation Rate (ν)	Daily probability of progressing from one exposed compartment to the next	0.584	[102, 114]
Infectious compartments (n)	Number of infectious compartments	3	Fit (See Section 3.6.1)
Recovery Rate (δ)	Daily probability of progressing from one infectious compartment to the next	0.3041	[102, 114]
Reproduction Number (R_0)	The expected total number of secondary infections from one infectious individual in a fully susceptible population	$0 - 3.1$	County R_0 estimates
Daily Reporting Rate (η)	The daily probability of an infectious individual being reported	Daily: $0.011 - 0.0224$ Overall: $10 - 20\%$	[50]
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	[114]

Table 3.1: 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.

with 95% confidence intervals spanning from 0 to 3.1 (Fig 3.6). Given this uncertainty, and that our primary aim is to demonstrate the risk assessment framework rather than provide accurate estimates of R_0 for Texas, we use these estimates to estimate relative county-level transmission risks (by scaling the county R_0 estimates from 0 to 1). In each simulation, we assume that a countys R_0 is the product of its relative risk and a chosen maximum R_0. For our case study, we assume a maximum county-level R_0 of 1.5 This is consistent with historical arbovirus activity in Texas (which has never sustained a large arbovirus epidemic) and demonstrates the particular utility of the approach in distinguishing outbreaks from epidemics around the epidemic threshold of $R_0 = 1$.

3.3.3 ZIKV Outbreak Simulation Model

Assuming mosquito-borne transmission as the main driver of epidemic dynamics, to transmit ZIKV, a mosquito must bite an infected human, the mosquito must get infected with the virus, and then the infected mosquito must bite a susceptible human. Rather than explicitly model the full transmission cycle, we aggregated the two-part cycle of ZIKV transmission (mosquito-to-human and human-to-mosquito) into a single exposure period where the individual has been infected by ZIKV, but not yet infectious, and do not explicitly model mosquitos. For the purposes of this study, we need only ensure that the model produces a realistic human-to-human generation time of ZIKV transmission, and the simpler model is more flexible to disease transmission

pathways. We fit the generation time of the ZIKV model to early ZIKV Epidemiological estimates, with further fitting details described in section 3.6.1.

The resultant model thus follows a Susceptible-Exposed-Infectious-Recovered (SEIR) transmission process stemming from a single ZIKV infection using a Markov branching process model (Fig 3.7). The temporal evolution of the compartments is governed by daily probabilities of infected individuals transitioning between disease states. New cases arise from importations or autochthonous transmission (Table 3.1). We treat days as discrete time steps, and the next disease state progression depends solely on the current state and the transition probabilities. We assume that infectious cases cause a Poisson distributed number of secondary cases per day (via human to mosquito to human transmission), but this assumption can be relaxed as more information regarding the distribution of secondary cases becomes available. We also assume infectious individuals are introduced daily according to a Poisson distributed number of cases around the importation rate. Furthermore, Infectious cases are categorized into reported and unreported cases according to a reporting rate. We assume that reporting rates approximately correspond to the percentage ($\sim 20\%$) of symptomatic ZIKV infections [**?**] 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

53

travel histories for all reported cases [32].

3.3.4 Simulations

For each county risk scenario, defined by an importation rate, transmission rate, and reporting rate, we ran 10,000 stochastic simulations. Each simulation began with one imported infectious case and terminated either when there were no individuals in either the Exposed or Infectious classes or the cumulative number of autochthonous infections reached 2,000. Thus the total outbreak time may differ across simulations. We held R_0 constant throughout each simulation, as we sought to model early outbreak dynamics over short periods (relative to the seasonality of transmission) following introduction. We classified simulations as either epidemics or self-limiting outbreaks; epidemics were simulations that fulfilled two criteria: reached 2,000 cumulative autochthonous infections and had a maximum daily prevalence (defined as the number of current infectious cases) exceeding 50 autochthonous cases (Fig 3.8 and 3.10). The second criterion distinguishes simulations resulting in large self-sustaining outbreaks (that achieve substantial peaks) from those that accumulate infections through a series of small, independent clusters (that fail to reach the daily prevalence threshold). The latter occurs occasionally under low R_0s and high importation rates scenarios.

To verify that our simulations do not aggregate cases from clear temporally separate clusters, we calculated the distribution of times between sequential cases (Fig 3.9). In our simulated epidemics, almost all sequentially

occurring cases occur within 14 days of each other, consistent with the CDCs threshold for identifying local transmission events (based on the estimated maximum duration of the ZIKV incubation period) [32].

3.3.5 Outbreak Analysis

Our stochastic framework allows us to provide multiple forms of 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, as defined above, as a function of the number of reported autochthonous cases. We call this epidemic risk. To solve for epidemic risk in a county following the xth reported autochthonous case, we first find all simulations 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 and only 50 of the 1,000 simulations ultimately fulfill the two epidemic criteria; the probability of detecting two cases in the county would be 10% and the estimated epidemic risk following two reported cases in that county would be 5%. This simple epidemic classification scheme rarely misclassifies 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, overlapping series of moderate outbreaks occasionally meet the two epidemic criteria. Under the highest importation rate we considered (0.3 cases/day),

55

only 1% of outbreaks were misclassified.

This method can be applied to evaluate universal triggers (like the recommended two-case trigger) or derive robust triggers based on 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 (fewer reported cases) as the policymaker threshold for action decreases, (e.g. 10% versus 30% threshold) and as the local transmission potential increases (e.g. $R_0 = 1.5$ versus $R_0 = 1.2$).

Variables ordered by importance
Total Amount of County Direct Spending on Traveling ($K)
Percentage Population holding Graduate or professional degree
Total Amount of Visitor Tax Receipts(Local) ($K)
County Male Population
Population Commuting to Work with Other Means
Max Temperature of Warmest Month
Percentage Population below Poverty Level
Precipitation of Wettest Quarter
Population without Health Insurance
Population holding Graduate or professional degree

Table 3.2: 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.

3.4 Results

ZIKV importation risk within Texas is predicted by variables reflecting urbanization, mobility patterns, and socioeconomic status table 3.2, and is concentrated in metropolitan counties of Texas (Fig 3.2A). In comparing the predictions of this model to out-of-sample data from April to September 2016, the model underestimated the statewide total number of importations (81 vs 151), but robustly predicted the relative importation rates between counties ($\beta = 0.97$, $R^2 = 0.74$, $p < 0.001$). The two highest risk counties–Harris, which includes Houston, and Travis, which includes Austin, have an estimated 27% and 10% chance of receiving the next imported Texas case respectively and contain international airports.

ZIKV transmission risk is concentrated in southeastern Texas (Fig 3.2B), partially overlapping with regions of high importation risk (Fig 3.2A). Our county-level estimates of R_0 range widely (from 0.8 to 3.1 for the highest-risk county), reflecting the uncertainty in socioeconomic and environmental drivers of ZIKV (Fig 3.6). We therefore analyzed the relative rather than absolute transmission risks. For purposes of demonstration, we assumed a plausible maximum county-level R_0 of 1.5, which closely followed our median estimates, and scaled the transmission risk for each county accordingly. The following risk analyses can be readily refined as we gain more precise and localized estimates of ZIKA transmission potential.

Wide ranges of outbreaks are possible under a single set of epidemiological conditions (Fig 3.3A). The relationship between what policymakers
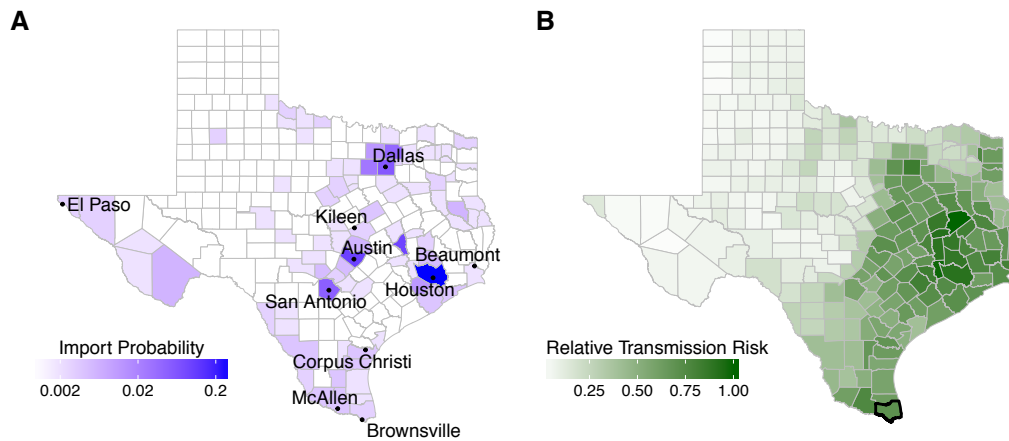
Figure 3.2: ZIKV importation and transmission risk estimates across Texas for August 2016. (A) Color indicates the probability that the next ZIKV import will occur in a given county for each of the 254 Texas counties. Probability is colored on a log scale. The 10 most populous cities in Texas are labeled. Houstons Harris County has 2.7 times greater chance than Austins Travis County of receiving the next imported case. (B) Estimated county-level transmission risk for ZIKV (See Fig 3.11 for seasonal differences). Harris county and Dallas County rank among the top 5 and top 10 for both importation and transmission risk respectively; counties in McAllen and Houston metropolitan area rank among the top 20. Bolded county border indicates counties with recorded local ZIKV transmission.

can observe (cumulative reported cases) and what they wish to know (current underlying disease prevalence) can be obscured by such uncertainty, and will depend critically on reporting rates (Fig 3.3B). Under a scenario estimated for Cameron County which experienced the only autochthonous ZIKV transmission in Texas and with a 20% reporting rate, ten linked and reported autochthonous cases correspond to 6 currently circulating cases with a 95% CI of 1-16 from inherent, early-stage outbreak stochasticity. From this wide
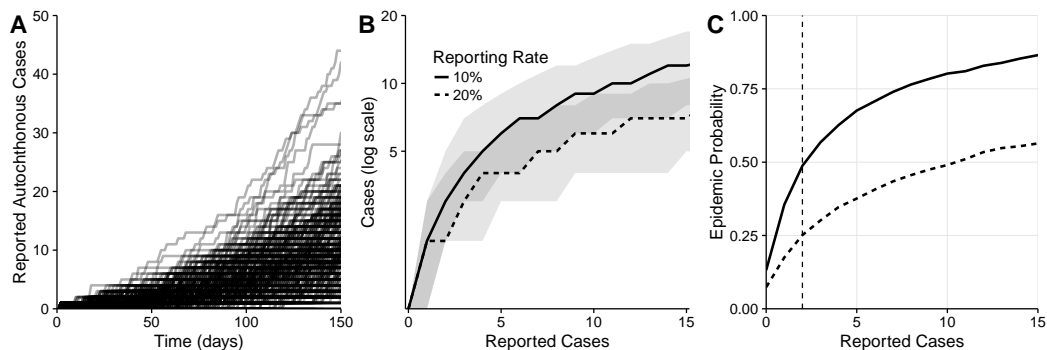
Figure 3.3: Real-time risk-assessment for ZIKV transmission. All figures are based on transmission and importation risks estimated for Cameron County, Texas. (A) Two thousand simulated outbreaks. (B) Total number of (current) autochthonous cases as a function of the cumulative reported autochthonous cases, under 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 for a low (solid) and high (dashed) reporting rate. Suppose a policy-maker plans to trigger a public health response as soon as a second case is reported (vertical line). Under a 10% reporting rate, this trigger would correspond to a 49% probability of an ensuing epidemic. Under a 20% reporting rate, the probability would be 25%.

range of outbreak trajectories, we can characterize time-varying epidemic risk as cases accumulate in a given county. We track the probability of epidemic expansion following each additional reported case in high and low reporting rate scenarios (Fig 3.3C).

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, this would correspond with a 49% probability of an epidemic if 10% of cases are reported, but only 25% if

the reporting rate is doubled.. Alternatively suppose a policy maker wishes to initiate an intervention when the chance of an epidemic exceeds 50%. In the low reporting rate scenario, they should act immediately following the third autochthonous reported case, but could wait until the eleventh case with the high reporting rate.
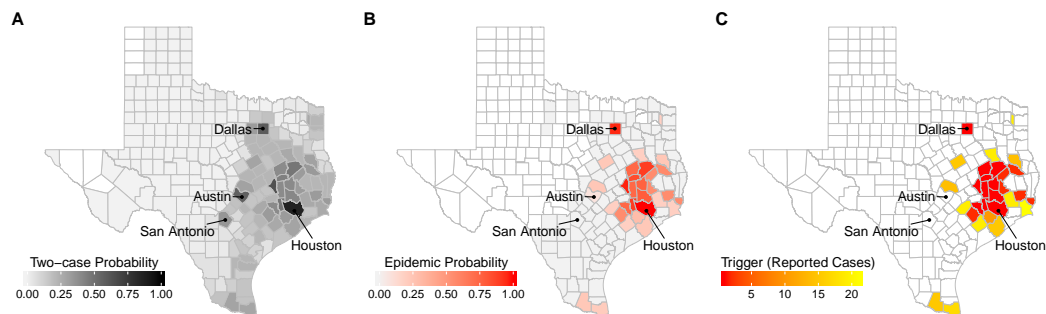
Figure 3.4: Texas county ZIKV risk assessment. (A) Probability of an outbreak with at least two reported autochthonous ZIKV cases. (B) The probability of epidemic expansion at the moment the second autochthonous ZIKV case is reported in a county. White counties never reach two reported cases across all 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.

To evaluate a universal intervention trigger of two reported autochthonous cases, we estimate both the probability of two reported cases in each county and the level of epidemic risk at the moment the trigger event occurs (second case reported). Assuming a baseline importation rate extrapolated from

60

importation levels in March 2016 to August 2016, county R_0 scaled from a maximum of 1.5, and a 20% reporting rate, only a minority of counties are likely to experience a trigger event (Fig 3.4A). While 247 of the 254 counties (97%) have non-zero probabilities of experiencing two reported autochthonous cases, only 86 counties have at least a 10% chance of such an event (assuming they experience at least one importation), with the remaining 168 counties having a median probability of 0.0038 (range 0.0005 to 0.087). Assuming that a second autochthonous case has indeed been reported, we find that the underlying epidemic risk varies widely among the 247 counties, 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 Harris County, correspond to a 99% chance of ongoing transmission that would proceed to epidemic proportions without intervention, with the rest of the Houston metropolitan also at relatively high risk ranging from 0 (Galveston) to 90% (Waller) (Fig 3.4B).

Given that a universal trigger may signal disparate levels of ZIKV risk, policy makers might seek to adapt their triggers to local conditions. Suppose a policymaker wishes to design triggers that indicate a 50% chance of an emerging epidemic (Fig 3.4C). Under the baseline importation and reporting rates, an estimated 31 of the 254 counties in Texas are expected to reach a 50% epidemic probability, with triggers ranging from one (Harris County) to 21 (Jefferson County) reported autochthonous cases, with a median of two cases. Counties who detect cases simply due to high importation rates do not

have triggers, and the magnitude of a trigger helps quantify a countys absolute risk for an epidemic as a function of the reported autochthonous cases.

3.5 Discussion

Our framework provides a data-driven approach to estimating ZIKV emergence risks from potentially sparse and biased surveillance data [20, 29]. By mapping observed cases to current and future risks, in the face of considerable uncertainty, the approach can also be used to design public health action plans and evaluate the utility of local versus regional triggers. We demonstrate its application across the 254 ecologically and demographically diverse counties of Texas, one of the two states that has sustained autochthonous ZIKV outbreaks [175, 176]. The approach requires local estimates of ZIKV importation and transmission rates. For the Texas analysis, we developed a novel model for estimating county-level ZIKV importation risk and applied published methods to estimate relative county-level transmission risks (Fig 3.2). We expect that most Texas counties are not at risk for a sustained ZIKV epidemic (Fig 3.4), and find that many of the highest risk counties lie in the southeastern region surrounding the Houston metropolitan area and the lower Rio Grande valley. However, R_0 estimates are uncertain, leaving the possibility that the R_0 could be as high as other high risk regions that sustained epidemics [7, 96, 151]. Our analysis is consistent with historic DENV and CHIKV outbreaks and correctly identifies Cameron county, the only Texas county to have reported local transmission, as a potential ZIKV hot-spot, especially when November estimates

are used [174] (Fig 3.11).

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. To assist Texas policymakers in interpreting the two-case trigger for intervention guidelines issued by the CDC [32], we used our framework to integrate importation and transmission risks and assess the likelihood and implication of a two-case event 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. Given this variation, rather than implement a universal trigger, which may correspond to 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. We found close agreement between the recommended two-case trigger and our epidemic derived triggers based on a 50% probability of expansion. Of the 30 counties with derived triggers, the median trigger was 2, ranging from one to 21 reported autochthonous cases. These findings apply only to the early, pre-epidemic phase of ZIKV in Texas, when importations

occur primarily via travel from affected regions outside the contiguous US.

These analyses highlight critical gaps in our understanding of ZIKV biology and epidemiology. The relative transmission risks among Texas counties appear fairly robust to these uncertainties, allowing us to identify high risk regions, including Cameron County in the Lower Rio Grande Valley. Public health agencies might therefore prioritize such counties for surveillance and interventions resources. Given the minimal incursions of DENV and CHIKV into Texas over that past eleven years since the first DENV outbreak in Cameron County, and the high number of importations into putative hotspot counties without autochthonous transmission, we suspect that, if anything, we may be underestimating the socioeconomic and behavioral impediments to ZIKV transmission in the contiguous US. Our analysis also reveals the significant impact of the reporting rate on the timeliness and precision of detection. If only a small fraction of cases are reported, the first few reported cases may correspond to an isolated introduction or a growing epidemic. In contrast, if most cases are reported, policymakers can wait longer for cases to accumulate to trigger interventions and have more confidence in their epidemiological assessments. ZIKV reporting rates are expected to remain low, because an estimated 80% of infections are asymptomatic, and DENV reporting rates have historically matched its asymptomatic proportion [**?**, 46]. 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 [96]; however estimates ranging from 1 to 10% have been reported during the ongoing epidemic in Columbia [**?**, 151]. 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, or active surveillance of regions with recent ZIKV cases. Our analysis assumes that all counties have the same case detection probabilities. However, only 40 of the 254 Texas counties maintain active mosquito surveillance and control programs, potentially leading to differences in case detection rates and surveillance efficacy throughout the state [162]. Thus, rapid estimation of the reporting rate using both traditional epidemiological data and new viral sequenced based methods [154] should be a high priority as they become available.

Our framework can 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.

3.6 Supplemental Information

3.6.1 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 [114]. 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 [110]. In such a framework one splits a compartment into multiple separate compartments (boxes), has individuals transition through these com-

partments, 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) [102]. Then, we fit the exposure period so that the combined duration of the infectious and exposure periods matched the empirical ZIKV generation time range [114], 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 [102]

County-level epidemiological estimates

Model inputs

General
- ZIKV transmission/disease parameters (Table S5)

County specific
- GDP
- Mosquito suitability
- Average monthly temperature

R_0

General
- State-wide importation rate (projected from historical estimates)

County specif c
- Ten environmental, socioeconomic and behavioral factors (Table S3)

importation rate

ZIKV outbreak simulations

10,000 stochastic branching process simulations for each county-risk scenario

Other parameters fit to published estimates
- ZIKV reporting rate
- ZIKV generation time
- Human infectious period

ZIKV risk analysis

County-level ZIKV risk estimation

Real-time ZIKV outbreak prevalence estimation

Development and interpretation of surveillance triggers

Figure 3.5: 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.
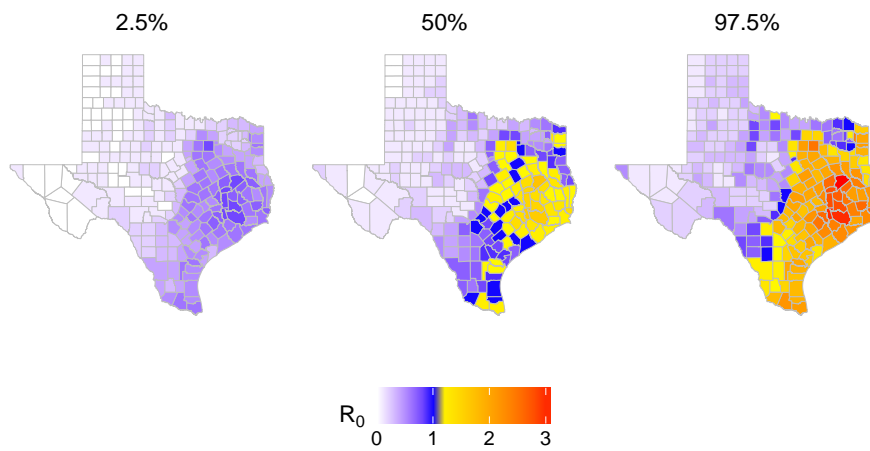
Figure 3.6: The 95% CI of R_0 Distributions for August. From left to right, the 2.5%, 50% and 97.5% quantile R_0 values for August. The range of absolute values spans 0.02-6.90. Given the considerable uncertainty in socioeconomic and environmental drivers of ZIKV, we analyzed relative rather than absolute transmission risks.

69
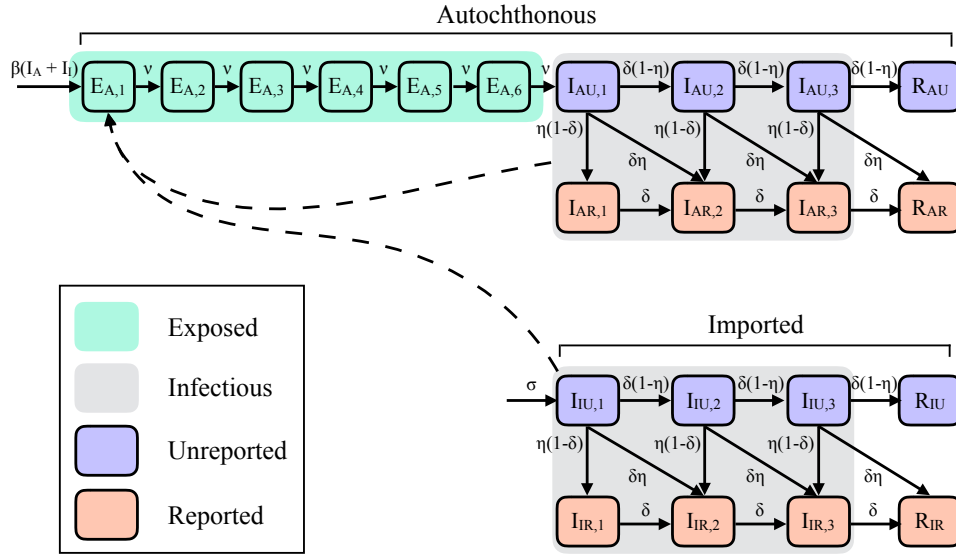
Figure 3.7: Diagram of ZIKV outbreak model. The model tracks disease progression, transmission, and reporting of both imported and autochthonous ZIKV cases. Individuals progress through compartments via a daily Markovian process, according to the solid arrows in the diagram. The *Exposed* and *Infectious* 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).
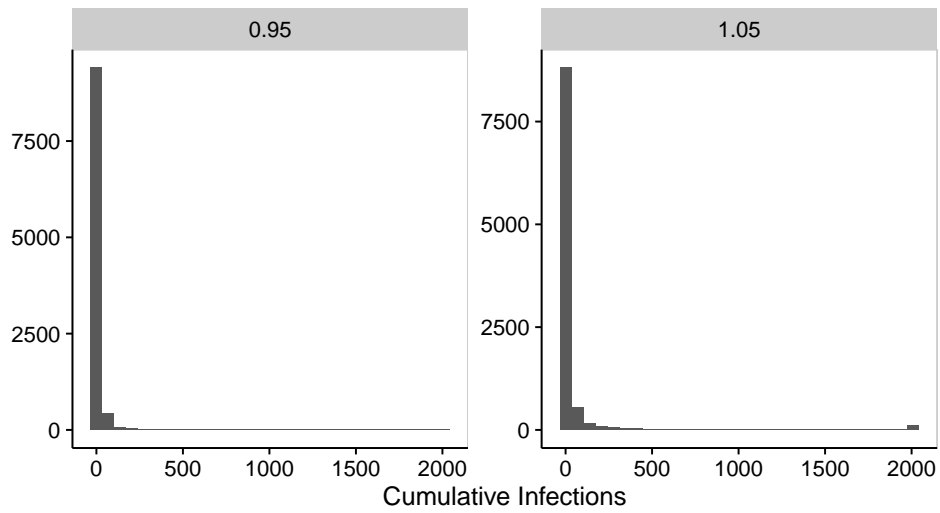
Figure 3.8: Determining outbreak simulation length. If outbreak simulations are too short, self-limiting outbreaks may reach the maximum number of infections due to stochasticity. We chose to run our simulations to 2,000 cumulative infections as it conservatively differentiated the large outbreaks of simulations with R_0 just below 1 (0.95) from the epidemics of those with R_0 just above 1 (1.05). We therefore chose to run our simulations until a maximum number of 2,000 infections.
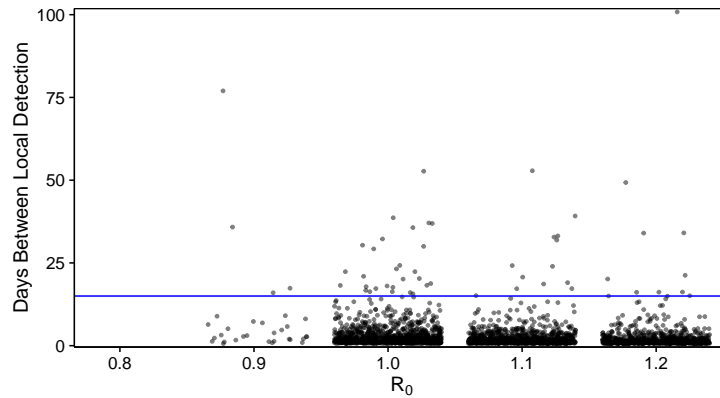
Figure 3.9: Time between detection of locally transmitted cases during epidemics. Across a range of R_0 values with an importation rate 0.1 cases/day, we plot the time between detection events of autochthonous cases for simulations out of the 10,000 trials in which epidemics occurred (black dots). The 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.
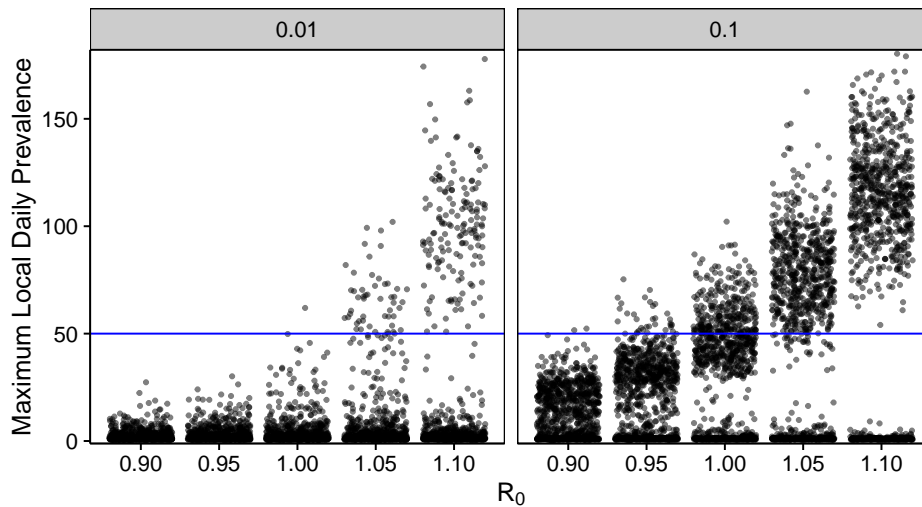
Figure 3.10: Selecting daily prevalence threshold for distinguishing self-limiting outbreaks from epidemics. Across a range of R_0 values, we plot the maximum daily total autochthonous infectious individuals for 1,000 of our 10,000 trials (black dots). The blue line indicates the threshold (50) selected 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.
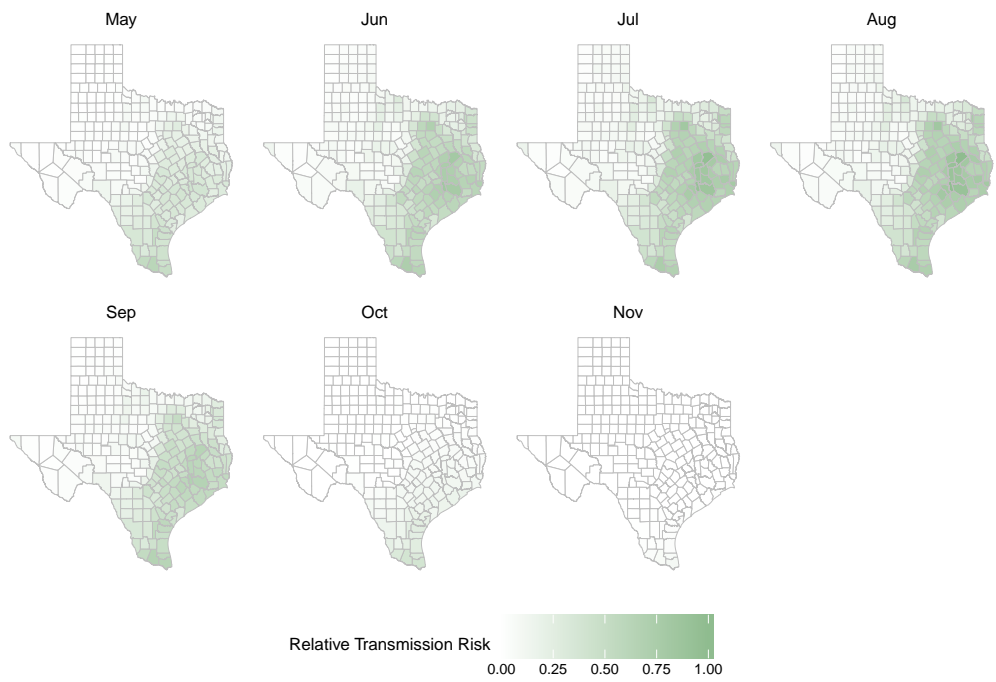
Figure 3.11: Monthly R_0 estimates based on seasonal changes in the temperature-dependent extrinsic incubation period of ZIKV in *Ae. aegypti* and the mosquito mortality rate of *Ae. aegypti*

.

Chapter 4: Downgrading disease transmission risk estimates using terminal importations

4.1 Abstract

As emerging and re-emerging infectious arboviruses like dengue, chikungunya, and Zika threaten new populations worldwide, officials scramble to assess local severity and transmissibility, with little to no epidemiological history to draw upon. Indirect estimates of risk from vector habitat suitability maps are prone to great uncertainty, while direct estimates from epidemiological data are only possible after cases accumulate and, given environmental constraints on arbovirus transmission, cannot be widely generalized beyond the focal region. Combining these complimentary methods, we use disease importation and transmission data to improve the accuracy and precision of a priori ecological risk estimates. We demonstrate this approach by estimating the spatiotemporal risks of Zika transmission throughout Texas, a high-risk region in the southern United States. Our estimates are, on average, 80% lower than published ecological estimates–with only six of 254 Texas counties deemed capable of sustaining a Zika epidemic–and they are consistent with the number of autochthonous cases detected in 2017. Real-time updating of prior risk estimates as importations and outbreaks arise can thereby provide critical, early insight into local transmission risks as emerging arboviruses expand

their global reach.

4.2 Introduction

The explosive emergence of Ebola in West Africa in 2014 and Zika in the Americas in 2016 caught the global health community by surprise. Officials scrambled not only to control the diseases at their source but also to anticipate and rapidly contain global transmission via infected travelers [69, 179]. The rate at which a newly introduced infectious disease spreads can vary enormously, depending on the physical and social environment. For example, serological surveys of dengue virus (DENV) exposure on either side of the Texas-Mexico border indicated far higher DENV exposure in the Mexican community despite virtually identical climatic conditions and even higher mosquito abundance in the Texan community [148].

Epidemiological risk assessment–estimating the severity and transmissibility of a threatening disease–can be vital to successful mitigation with limited resources. Historical outbreak data can provide invaluable insight into future epidemic risk. However, for a disease that has yet to arrive or has just begun to spread, we are forced to borrow epidemiological data from other populations or related diseases, or to indirectly assess risk based on environmental suitability. For example, as the first importations of Zika virus (ZIKV) arrived in the US in 2016, early attempts to determine the likelihood and rate of local transmission relied primarily on dengue epidemiological data from regions with markedly different climatic and socioeconomic conditions [7, 25, 28].

These ecological risk assessments provide information regarding the reproduction number of a disease (R_0)—the expected number of secondary human infections resulting from a single human infection—which provides a meaningful and predictive measure of local epidemiological risk. In a naive population, R_0 indicates whether importations can potentially ignite local epidemics; if so, it also provides insight into the probability, magnitude, and speed of spread [112, 186]. However, ecological estimates often carry considerable uncertainty stemming from model parameterization and regional extrapolation, and suggest a wide-range of possible epidemic outcomes, from terminal importations to stuttering chains of transmission to full blown epidemics [24, 28, 53, 77]. Once an outbreak is underway, we can use early case data to directly estimate R_0 [20, 21, 75]. For arboviruses with environmentally constrained spatial heterogeneity, we cannot easily extrapolate such case-based estimates from one region to others.

Here, we introduce a method for estimating the R_0 of an emerging arbovirus prior to a large outbreak in populations that face the ongoing threat of infected travelers from affected regions. This approach was motivated by recent introductions of ZIKV into the continental US. As hundreds of cases arrived from affected regions throughout the Americas, officials sought to estimate risks of autochthonous (local) transmission and identify high risk regions in the southern US. However, given the novelty of ZIKV and the large proportion of ZIKV cases that go undetected, early ecological estimates had high uncertainty [7, 25, 50, 72, 118]. Our method combines indirect and direct es-

77

timation methods to reduce such uncertainty and increase accuracy at high spatiotemporal resolution. We first build prior ecological estimates of local R_0 and then harness real-time importation data—cases that arrive in a naive location with or without subsequently infecting others—to update the estimates, while explicitly modeling case reporting uncertainty. As a case study, we use the almost complete absence of secondary transmission following 298 importations of ZIKV into the state of Texas in 2016 and 2017 to reduce and narrow local estimates of R_0.

4.3   Methods

We used a two-step procedure to estimate the monthly R_0 for each of the 254 Texas counties (hereafter county-month R_0): (1) estimate a priori county-month R_0 distributions using published ecological models of ZIKV transmission [7, 28], and (2) using these as Bayesian priors, generate posterior R_0 distributions based on reported importations and subsequent local transmission.

4.3.1   Data

We analyzed all ZIKV importations into Texas from January 2016 to September of 2017, including the county and notification date; county-level purchasing power parity (PPP) in US dollars [137]; daily temperature data at a 5 km x 5 km resolution for 2016-2017 and historical averages from 1960-1990 [78, 136]. For each county and month, we averaged daily temperatures across

all 5 km x 5 km grid cells whose center fell within the county; we aggregated 5 km x 5 km mosquito (Aedes aegypti) occurrence probabilities similarly [93]. Data available `https://dx.doi.org/10.18738/T8/HYZ53B`.

In all, six mosquito-borne, autochthonous cases of ZIKV were reported in Texas in 2016 and two were reported in 2017 [178]. For updating R_0 estimates, we analyzed 2016 data and assumed that two autochthonous cases were detected in Cameron County–one in November and one in December 2016; we excluded four nearby cases discovered during the November follow-up investigation, because our model does not incorporate active surveillance. As sensitivity analyses, we re-estimated R_0 assuming that no cases were detected and that all six cases were detected (Fig 4.12). For validating our estimates, we analyzed 2017 data and considered only one of the two reported autochthonous cases, as the second case occurred outside the timeline of our 2017 importation data.

4.3.2 A priori county-month R_0 estimates

Following Perkins *et al* [7], we estimated R_0 using the Ross-Macdonald temperature-dependent formulation:

$$R_0(T) = m\frac{bca^2e^{-\mu(T)n(T)}}{\mu(T)r} \tag{4.1}$$

with parameters as defined in Table 4.1. We calculated relative abundance of the ZIKV vector based on Ae. aegypti occurrence probabilities as $-ln(1-\rho)$, where ρ is the occurrence probability, and we interpret this as a

relative (rather than absolute) abundance, which is sufficient for our R_0 estimation [7]. We derived *a priori* county-month R_0 distributions by drawing 1,000 Monte Carlo samples from each underlying parameter distribution, with the appropriate county and month data. Finally, we fit gamma distributions to each probability distribution for use as an informative priors.

Table 4.1: Parameters of prior R_0 estimates.

Parameter	Description	Distribution	Value (CI)	Citation
b	Mosquito-to-human transmission probability	Constant	0.4	[4]
$\frac{c}{r}$	Human-to-mosquito transmission probability times the duration of human infectiousness	Constant	3.5	[158]
a	Mosquito biting rate	Constant	0.67	[147]
$\mu(T)$	Mosquito daily mortality rate	Non-parametric GAM	0.115[1]	[22, 128]
$n(T)$	Extrinsic incubation period in mosquitoes	Exponential	6.1 (3.4, 9.9)[2]	[33]
m	Economic mosquito-human contact factor	Monotonic decreasing SCAM	0.67	[7]

1. Fit to data from mark-recapture study occurring between 20-34°C

2. At 30°C

4.3.3 Autochthonous transmission likelihood

Following [21], we developed a likelihood function describing the expected outbreak size following an importation. We assumed that the secondary case distribution for each infected is negative binomial with mean R_0 and dispersion parameter, k. Assuming all cases are detected, the probability of an outbreak of chain size, j, is given by:

$$s(j, R_0, k) = \frac{\Gamma(kj + j - 1)}{\Gamma(kj)\Gamma(j + 1)} \frac{(R_0/k)^{j-1}}{(1 + (R_0/k))^{kj+j-1}} \tag{4.2}$$

where $\Gamma(n) = (n - 1)!$. However, not all cases are detected and the imported index case is always detected and correctly classified as an importation, so the probability of detecting a chain of size, j, from a given importation is given by:

$$s^*(j, R_0, k) = \sum_{l=j}^{\infty} s(l, R_0, k) \cdot \binom{l - 1}{j - 1} \cdot p_d^{j-1} \cdot (1 - p_d)^{l-j} \tag{4.3}$$

where p_d is the case detection probability. Importantly, this allows for local, undetected cases. We take the product of all likelihoods for each imported case as

$$\mathcal{L}(\vec{O}|\alpha, \vec{R_0}, k, p_d) = \prod_{i=1}^{\text{length}(\vec{O})} s^*(O_i, \alpha R_{0\gamma_i, \omega_i}, k, p_d) \tag{4.4}$$

where $\vec{O}$, contains the observed outbreak sizes for each importation (terminal importations have an outbreak size of one), $R_{0\gamma_i, \omega_i}$ denotes the county(γ)-month(ω) R_0 for the location and time that the importation occurred, and α is a statewide scaling factor applied to each $R_{0\gamma_i, \omega_i}$. The introduction of the state-wide scaling factor allows for localized importations to inform statewide estimates, but assumes that biases in the a priori R_0 estimation procedure are constant across counties and months. Details of simulations

and validation of the likelihood can be found in supplemental section 4.6.1 and Fig 4.6.

4.3.4 Estimating the dispersion parameter

The negative binomial dispersion parameter governs the variability in secondary cases following each importation, with values near zero meaning that most importations yield few or no cases while a few superspreaders produce many. We assume that ZIKV secondary case distributions resemble that of dengue virus (DENV) [141]. Padmanabha et al. describe the relationship between regional R_0 and the percentage of DENV cases generating over 20 secondary infections (p_{20}), as $R_0 = 0.63 \times 100(p_{20}) + 0.58$. We assumed that $p_{20} = 1 \times 10^{-8}$ for $R_0 < 0.58$, and found that a single dispersion parameter captures this relationship for all R_0 values and thus used $k = 0.12$ for all analyses (Fig 4.7).

4.3.5 Updating posterior R_0 estimates

We estimated posterior distributions for α, and each county-month R_0 for each day with a new importation between January 2016 and January 2017. We assumed a uniform prior for α of $U \sim (0, 2)$, and used a blocked Gibbs sampling algorithm of MCMC. For each MCMC step we provide the detected imported cases to date and propose each county-month R_0, a single α, and a p_d. County-month R_0 proposals were normally distributed around the previous sample with standard deviation of 0.1, α proposals were distributed $U \sim (0, 2)$,

and we used a previously estimated distribution for the reporting rate, $p_d \sim N(5.74\%, sd = 1.49\%)$, which we assumed to not vary spatiotemporally [199]. We used the Metropolis-Hastings probability to accept or reject proposals. Our chains consisted of 200,000 samples with a burn-in duration of 100,000; thinning every 10 steps. Further algorithmic details and code are available on Github (`https://github.com/sjfox/rnot_updater`).

4.3.6 Validating posterior county-month R_0 estimates

We derived the expected number of autochthonous cases from the importations data through September of 2017 (at that time, the most recent importation was detected in mid-May) and compared the estimates to the actual reported autochthonous cases. We integrated uncertainty into our estimates by sampling from the posterior county-month R_0 distributions and simulating outbreaks accordingly (full details in section 4.6.2).

4.4 Results

4.4.1 Importation-based updates of transmission risk

Hypothetically, suppose that the first 15 imported cases of Zika into Texas arrived in August into Harris County (which contains Houston) without any detected autochthonous transmission. Prior to these importations, environmental suitability models yielded a relatively high local risk estimate with median Harris county R_0 above the epidemic threshold of one (Figure 4.1A - dark grey). The lack of secondary cases following all 15 importations

suggests that R_0 may be lower. Indeed, our updated estimates suggest that the Harris county R_0 is likely below one (Fig 4.1A - light grey). Our method leverages such county-level importation data to update R_0 estimates throughout the state (via a scaling factor), based on the assumption that any *a priori* biases will be similar across counties (Fig 4.1B).
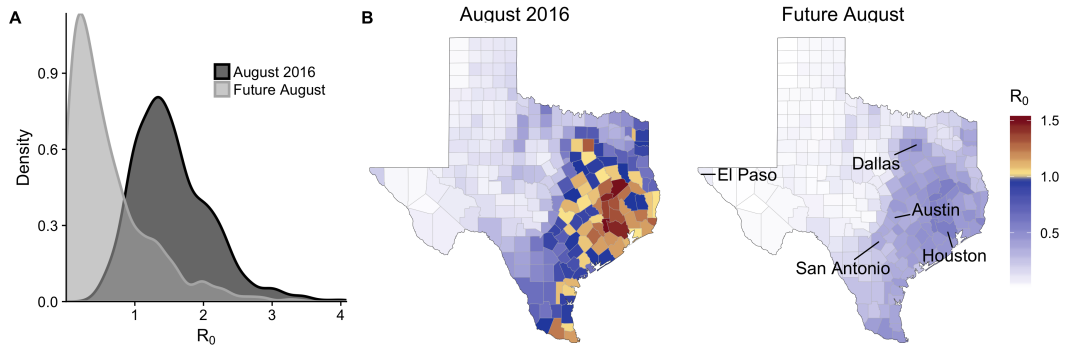
Figure 4.1: R_0 updating using importation data. Consider a hypothetical scenario in which the first 15 terminal ZIKV importations into Texas arrive in Harris county (which includes Houston) during August 2016. (A) Estimated Harris county R_0 for August 2016 a priori (dark grey) and after accounting for the 15 (light grey) terminal importations (Future August). (B) Median R_0 estimates for August before (August 2016) and following (Future August) the importation-based update.

4.4.2 Baseline importation and transmission risks in Texas

Prior to making importation-based updates, our initial median estimates of R_0 across Texas 254 counties in 2016 range from approximately 0 to 1.5 throughout the year with July and August having the highest transmission risk (Fig 4.2A). Throughout the manuscript, we conduct a one-sided test at a 1% significance level and thus consider counties with 99 percentiles (upper

bounds) that include one to be at risk for an epidemic ($R_0 > 1$). Initial upper bound estimates reach as high as three, and 119 (47%) of Texas counties are expected to be at risk of a local outbreak in at least one month of the year (Fig 4.2A, 4.8). When we considered historic average temperatures rather than 2016 temperatures, the projected 2017 risks were consistently lower, with the largest differences occurring during the unseasonably warm 2017 winter (Fig 4.9). Case importations peaked in July, August, and September of 2016, with 164 (55%) of the 298 total 2016 importations arriving then (Fig 4.2B). The few detected autochthonous cases occurred in November and December, when expected risk was relatively low but not negligible.

4.4.3   Updated transmission risks in Texas

Based on all importations and autochthonous cases that occurred in Texas prior to January 2017, we estimate that all Texas counties have a median posterior R_0 below one (Fig 4.3). Median estimates range from 0 to 0.29; upper-bound estimates range from 0 to 1.12, with only six (5%) of the original 119 high-risk counties maintaining epidemic potential (Fig 4.10). When we assume historic averages rather than 2016 temperatures, we obtain similar results (Fig 4.11).

In a sensitivity analysis that assumes 20 times more undetected importations, we found that the estimated risks decreased further (Fig 4.12). We also varied the number of detected autochthonous cases in November: as they decrease from one to zero, the estimated risks decrease considerably; as they

85
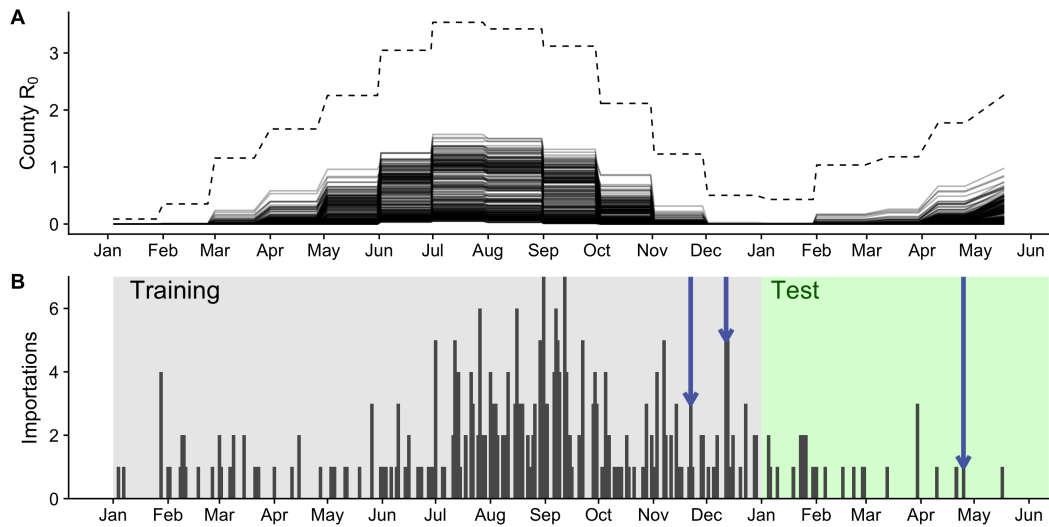
Figure 4.2: Texas importations and baseline transmission risk estimates for 2016-17. (A) Initial ZIKA R_0 estimates using ecological risk models parameterized with actual 2016-2017 temperatures. Each solid line shows median values for one of Texas 254 counties. Dashed line shows the highest upper bound (99th percentile) across all counties. (B) Daily ZIKV importations into Texas. Blue arrows indicate importations that produced detected autochthonous transmission; shading indicates training (2016) and testing (2017) periods.

increase to five, estimated risks increase, with 83 counties becoming at risk for a local outbreak (Fig 4.12).

Importation events had variable impacts on the posterior estimates, depending on their timing and location (Fig 4.4). Terminal importations early in the year, when a priori R_0 estimates were low, had little effect; those arriving in the summer months, when high a priori R_0 estimates suggested that transmission should have occurred, led to sharp decreases and a shrinking confidence interval. By early November, the median α decreased from 1.0 to 0.06
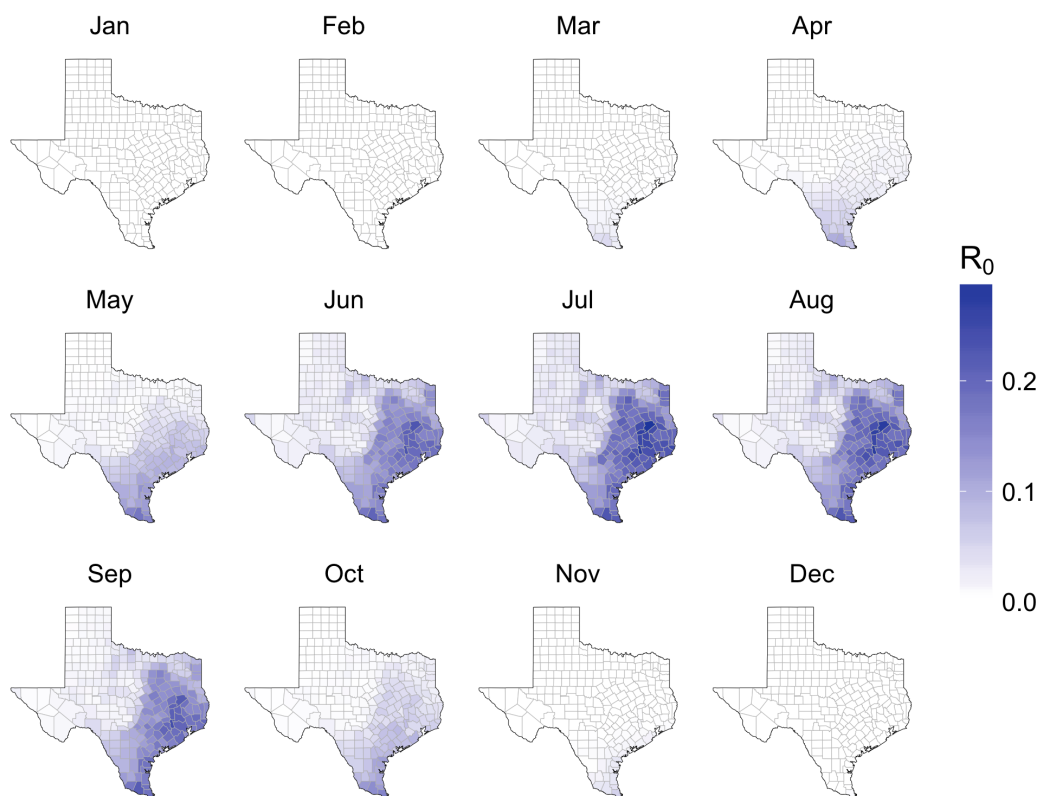
Figure 4.3: Posterior median county R_0 estimates for Texas, based on ZIKV importations through January 2017. This assumes that all importations were terminal except for a two autochthonous cases detected in Cameron County in late 2016.

with a narrow 95% CI of 0.002-0.30. However, the two secondary transmission events detected in November and December increase R_0 estimates and widen the confidence intervals. Incorporating all data up to January 2017, our best estimate is that R_0 values across the state are roughly one fifth the original estimates (median: 0.19, 95% CI: 0.05-0.48).

4.4.4 Expected autochthonous transmission in Texas

We use transmission risk estimates based on importations through December 2016 to estimate the number of autochthonous cases we would expect to detect in Texas in 2017. Assuming first that only the reported importations occurred in 2017 (26 total), we estimate that there should have been 0.08 (95%CI: 0-1) detected autochthonous cases in the state; assuming that many importations went undetected, according to the reporting probability $(\frac{26}{p_d} \approx 453)$, we estimate 1.3 (95% CI: 0-7) detected autochthonous cases. Both of these estimates are consistent with the single autochthonous case detected in Texas in 2017, though our results best fit a scenario with many undetected importations (Fig 4.5).

4.5 Discussion

The global expansion of ZIKV was declared a Public Health Emergency of International Concern in February 2016, and caused more than 565,000 confirmed or probable cases and over 3,352 documented cases of congenital Zika syndrome. Although it is receding in most regions of the world, ecological risk assessments suggest that previously unaffected or minimally affected areas may remain at risk for future emergence, including parts of Asia and South America [12, 44, 163]. Differentiating regions that can sustain a ZIKV epidemic ($R_0 > 1$) from those that cannot is vital to effective planning and resource allocation for future preparedness plans. To address this challenge, we have developed a simple method for refining uncertain risk assessments with readily

available data on disease importations.

We applied the method to update ZIKV R_0 estimates for each of the 254 counties in Texas, and found that only six counties have non-negligible probabilities of sustained local transmission. This is a substantial downgrade in expected risk, given that 43% of the 254 counties were previously thought to be vulnerable to ZIKV outbreaks [28]. These estimates suggest that there should have been roughly one detected case of locally acquired ZIKV between January and September of 2017, closely corresponding to the single transmission event actually detected in Cameron County in July 2017 (Fig 4.5). Our sensitivity analysis suggests that, if our assumptions about case-reporting in November are too liberal, 77 additional counties have non-negligible but low risks of summer outbreaks. Given comparable importation and climatic data, this approach could readily update ZIKV transmission risk estimates for all counties in the continental US and elsewhere.

Our estimation method relies on several simplifying assumptions. We assumed that the shape of the secondary case distribution resembles that of dengue. Although we have no evidence to the contrary, this should be updated as ZIKV-specific estimates become available [141]. We also assumed that transmission is equally likely from imported and locally acquired cases. Imported cases may be less infectious than locally acquired cases for two reasons, leading us to underestimate local transmission risks. First, they may be more likely to receive care that limits transmission, although most ZIKV cases are inapparent or mild and do not require medical care [50], and second their

89

local infectious periods may be shorter than those of autochthonous cases. Next, we treat all importations as independent. However, spatiotemporal heterogeneity in case detection probabilities or clustering of cases (e.g., testing of travel companions) could bias risk estimates. Furthermore, when secondary clusters are detected, we assume they share a transmission tree stemming from a single detected importation. In fact, the low ZIKV detection rate suggests that both primary importations and secondary cases are likely to be missed. If the detection rates are roughly similar, our results hold. When we assume, in sensitivity analysis, that importations are detected at higher rates than secondary cases, then the resulting risk estimates will be higher; when we assume the reverse, they are lower. The additional assumption, that clusters are epidemiologically connected, seems reasonable for the small contained outbreaks detected in Texas, but may not be appropriate for importation-fueled arbovirus outbreaks in Florida, for example. In such cases, molecular data might enable estimation of transmission clusters [70, 182]. We also rely on informative Bayesian priors and a statewide scaling factor, which allows us to use local importations to inform risk estimates elsewhere, but implies that our prior county-month transmission risk estimates are correct relative to each other. Given additional importation data, we could potentially estimate each county-month R_0 independently. Finally, we do not consider possibility of sexual transmission of ZIKV. While sexual transmission has occurred and may be important for specific populations [9], we assumed that mosquito-borne transmission is the dominant mode of infection.

During the height of the ZIKV threat, public health agencies in the US rapidly implemented both preventative measures (e.g., vector control and educational campaigns) and response measures (e.g. laboratory testing and epidemic trigger plans), particularly in high risk southern states. Decision makers sought to identify and narrow the spatiotemporal scope of outbreak risk to enable targeted responses, efficiently allocate resources, and avoid false alarms [32, 72]. Our method facilitates such rapid, real-time geographic risk estimation from typical early outbreak data, and suggests that only 3% of the Texas population is at risk for a local outbreak. Critically, we can conclude neither that all initial ecological risk assessments for ZIKV will overestimate risk, although this seems to be the case for ZIKV in Texas, nor that public health preparations and interventions for ZIKV are no longer necessary in Texas or the southern US. Rather, our results suggest that sustained ZIKV outbreaks are unlikely, but not impossible, and provide more robust and localized estimates of ZIKV risk that can inform more targeted surveillance and reactions to future ZIKV importations.

This framework is novel in its integration of *a priori* ecological transmission risk estimates with updating directly from real-time case reports [20, 21]. It thereby provides increasingly accurate and precise risk assessments to support public health decision making. This approach can be generalized to update R_0 estimates from importation data, regardless of the *a priori* method of estimation. For example, a new approach combining epidemiological and molecular analyses suggests that transmission risk in Florida is subcritical (i.e.,

91

$R_0 < 1$) [47, 70]. Given that Florida experienced thousands of introductions, only a few of which sparked large outbreaks, coupling such outbreak-driven estimation with our terminal importation method may provide a powerful real-time risk assessment framework for exploiting all available data. This method resembles those used to assess disease transmission risk during elimination efforts, including malaria in non-endemic regions [39]. The key innovation is that, by starting with ecological suitability maps, we simultaneously identify important transmission hotspots and leverage case data from one region to inform risk estimates elsewhere.

We presented a simple and rational method for continuously updating transmission risk estimates for populations experiencing infectious disease importations, with or without secondary transmission. As we demonstrated for ZIKV in Texas, large numbers of terminal importations can profoundly lower both estimated risks of transmission and uncertainty in prior estimates, particularly those derived from ecological suitability or other models that borrow inputs from related pathogens in other parts of the world. Expanding our model globally could drastically reduce the expected number of populations at risk [24, 25, 118]. Although the threat of ZIKV emergence in the continental US motivated this study, this new framework can be widely applied to improve transmission risk assessments when a disease newly threatens a population via regular introductions with minimal secondary transmission. For example, importation-fueled MERS-CoV transmission risk, measles transmission in vaccinated populations, or highly pathogenic avian influenza [41, 105, 194].

4.6 Supplemental Information

4.6.1 Outbreak Simulation

We checked our analytical calculations against simulations to ensure consistency. The outbreak simulations consist of two steps: (1) the outbreak transmission process and (2) the observation process. We conduct three simulation scenarios to test against analytical results. The first two simply confirm prior results, and the third confirms our analytical formulation. For each scenario, we obtain 10,000 simulations, and summarize the resultant outbreaks to obtain a probability mass function for the observed outbreak size. We then compare these simulated probability mass functions with the analytical calculations presented in the methods, and find close agreement (Fig S1).

1. Perfect observation. In this simulation, we assume that all cases are detected. For each simulation we begin with an initial index case. That case, and all subsequent cases, infect individuals according to a negative binomial distribution of mean, R_0, and dispersion parameter $k = 0.12$. We continue the simulation until there are no more newly infected individuals. We only focus on $R_0 < 1$ for simulation purposes, so no outbreaks grow forever.

2. Imperfect observation. In this simulation, we add the observation process to the transmission process. We therefore simulate outbreaks according to the perfect observation process described above, and then simulate the reporting process. For this, we find the total number of detected cases

from the chain according to simulating a binomial detection process with probability of success equal to the reporting rate (0.0574 in this case) and total possible cases equal to the size of the chain from the transmission chain (n).

3. Imperfect Import observation. This simulation is exactly the same as the imperfect observation simulation, except for the fact that we always detect the index case. So in this case we simulate a binomial detection process with probability of success equal to the reporting rate (0.0574 in this case) and total possible cases equal to the size of the transmission chain minus one ($n - 1$).

4.6.2 R_0 estimate validation

To validate our posterior R_0 distributions, we used them to estimate the expected number of autochthonous cases from the importations data through September of 2017 (at that time, the most recent importation was detected in mid-may) and compared the estimates to the actual reported autochthonous cases. We integrated uncertainty into our estimates as follows:

1. Draw a p_d from the reporting rate distribution.

2. Sum the number of importations occurring for each county-month, N,

3. Draw N, samples of the prior or posterior R_0 , distribution depending on which analysis is being conducted.

4. For each of the R_0, values, we simulated an outbreak stemming from a single importation where each case infects individuals according to a negative binomial distribution with mean of R_0, and dispersion parameter, $k = 0.12$. For each simulated outbreak, we simulate the detection process for the non-index cases as a binomial distribution with probability of success, p_d. We sum the detected cases for each of the N, outbreaks, to obtain, ν, the expected number of cases detected for that sample.

5. Repeat steps 1-4 10,000 times, saving ν

The distribution of ν obtained from the process described above can be compared with the true number of detected autochthonous cases from 2017 if we assume that all imported cases were reported. However, its likely that there were a number of imported cases that were missed by surveillance. Therefore, we also analyzed a scenario with increased importations. To do so we followed the same process outlined above, except for altering step 3 to draw $N * \left(\frac{1}{p_d}\right)$ samples to account for the missed cases rounding the resultant number to the nearest integer.

Supplemental Figures
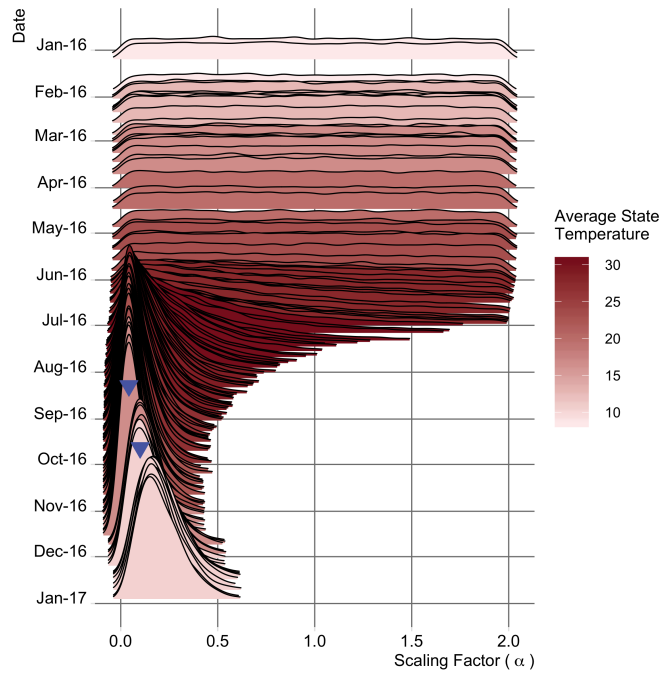
Figure 4.4: Evolving posterior distribution of statewide scaling factor for R_0. Zika importations, both with and without subsequent detected autochthonous transmission, provide insight into local transmission potential, via a statewide scaling factor, α. This shows the posterior distributions of α, for each day of 2016 that had at least one imported case. Median estimates reach a minimum in early November, just before the detected autochthonous transmission events (upside-down blue triangles). Red shading indicates the average statewide monthly temperature. Note: the scaling factor is never less than zero.
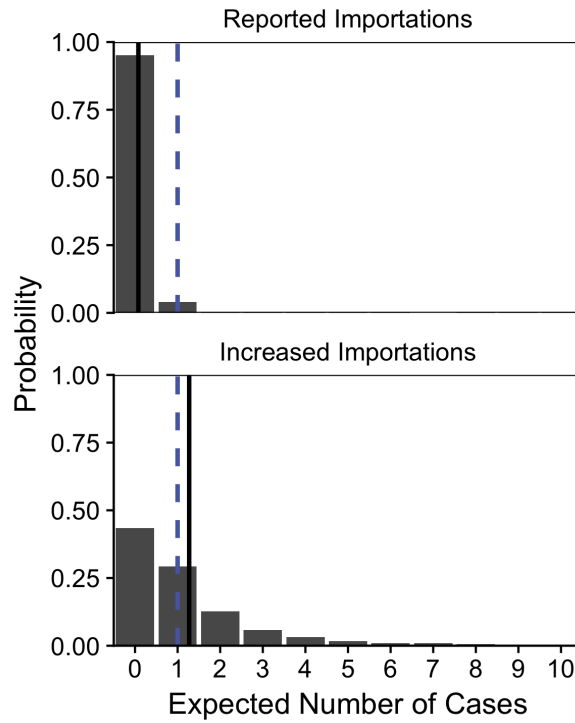
Figure 4.5: Expected autochthonous cases in 2017, assuming revised county R_0 estimates and reported importations through September 2017. The probability distributions are built from 10,000 simulations, each randomly drawing from the R_0 posterior distributions. The dashed blue line indicates the actual number of detected autochthonous cases in state (one), and the solid black lines indicate means for the baseline importation scenario, in which only the reported importations occurred (top) and the increased importation scenario, in which a large fraction of importations went undetected (bottom).
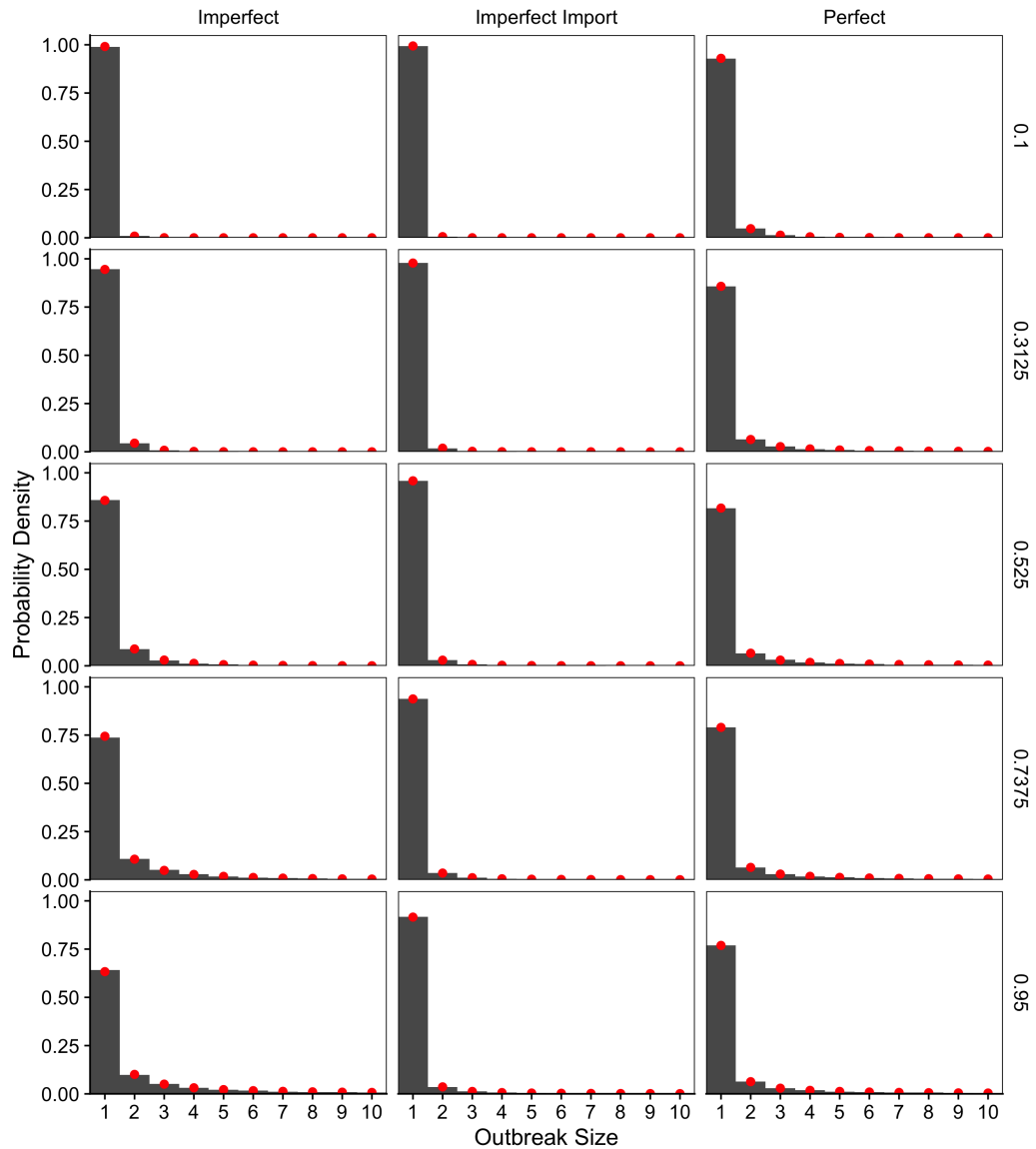
Figure 4.6: Comparison between analytical likelihood predictions and simulations. We compare the probability mass functions for the outbreak sizes for our simulations (bars) with the analytical expectation (red dots). Rows demonstrate four different transmission risk scenarios (R_0), and columns describe three different scenarios: (1) where every case within a transmission chain is detected (Perfect), (2) where all cases are detected independently with a specific reporting rate (Imperfect), and (3) where all cases are detected independently with a specific reporting rate except for the index case which is always detected (Imperfect Import). The Imperfect Import probability mass function is the one used for all analyses in this article. All simulations are completed with a reporting probability of 0.0574, and $k = 0.12$.
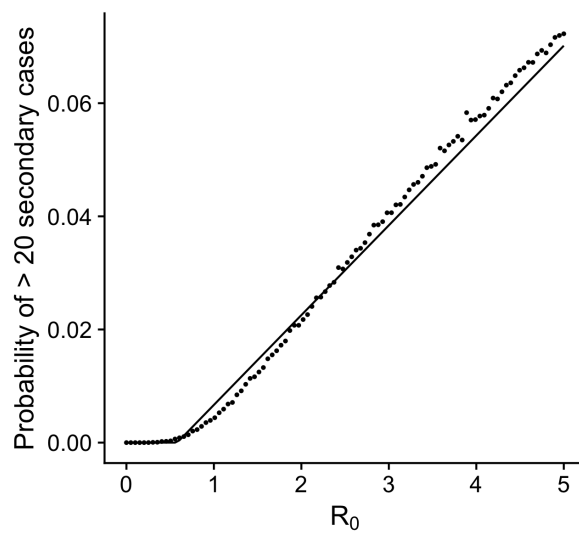
Figure 4.7: Match between estimated and assumed dispersion parameter. Probability of a single importation generating 20 secondary infections from (29) (Line), or using our assumed dispersion parameter and negative binomial distribution (Points).
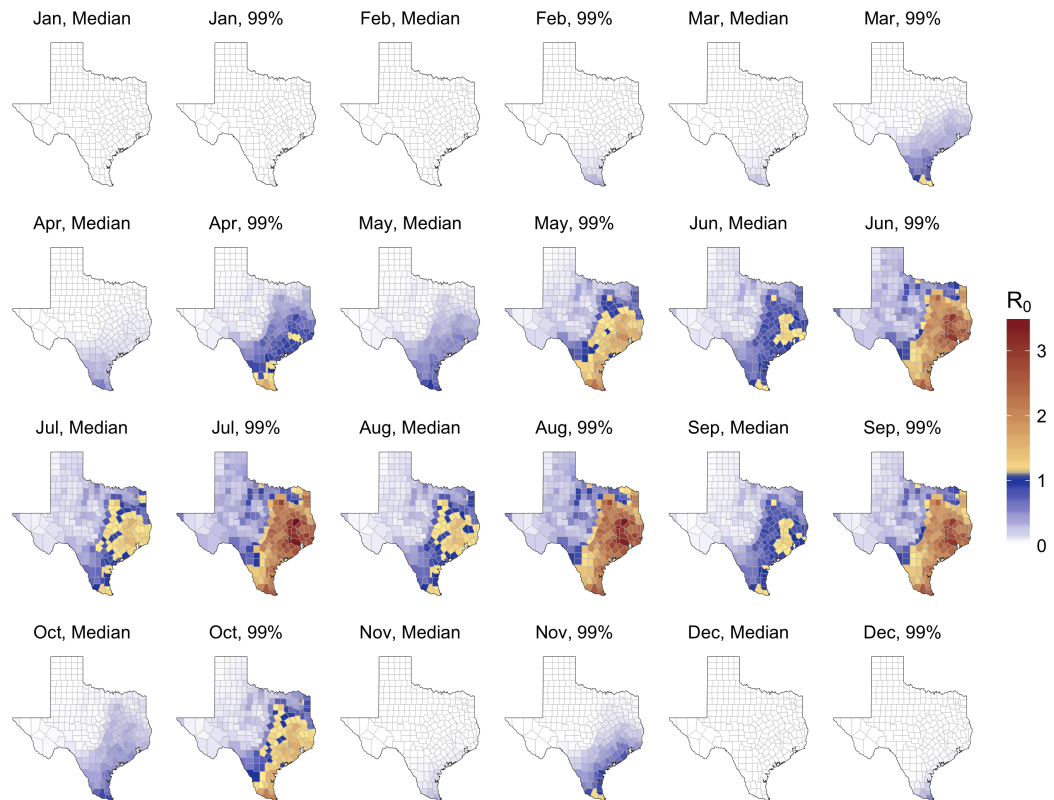
Figure 4.8: Prior R_0 estimates for Texas counties for each month. Median and 99 percentiles are shown for each County. Fill color indicates the estimated Median or 99 percentile estimate for that county for the given month, with counties showing yellow or red indicating their R_0 is above one (labels). Estimates are made for each month based on the average monthly temperature for 2016.
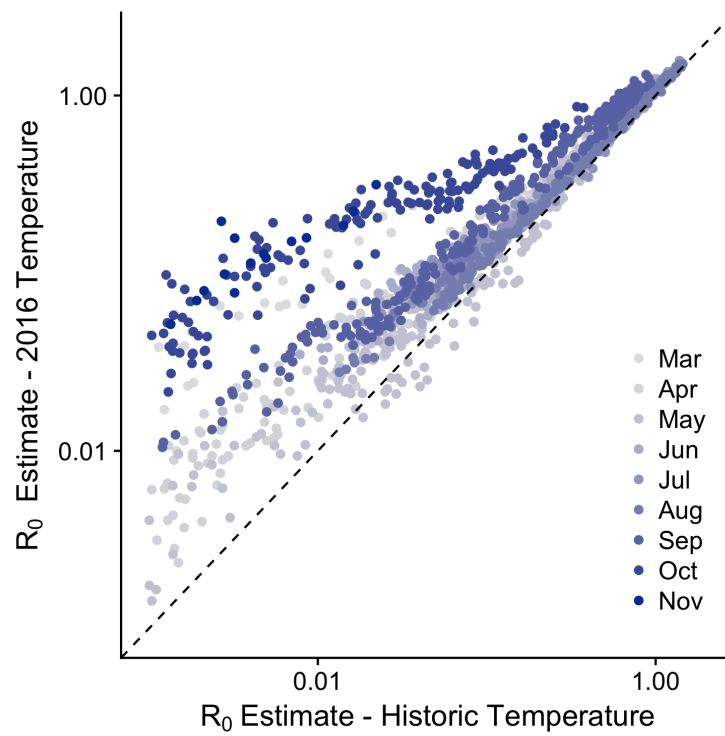
Figure 4.9: Comparison between prior R_0 estimates using the historic average temperature for months versus using the actual temperature from the months in 2016.
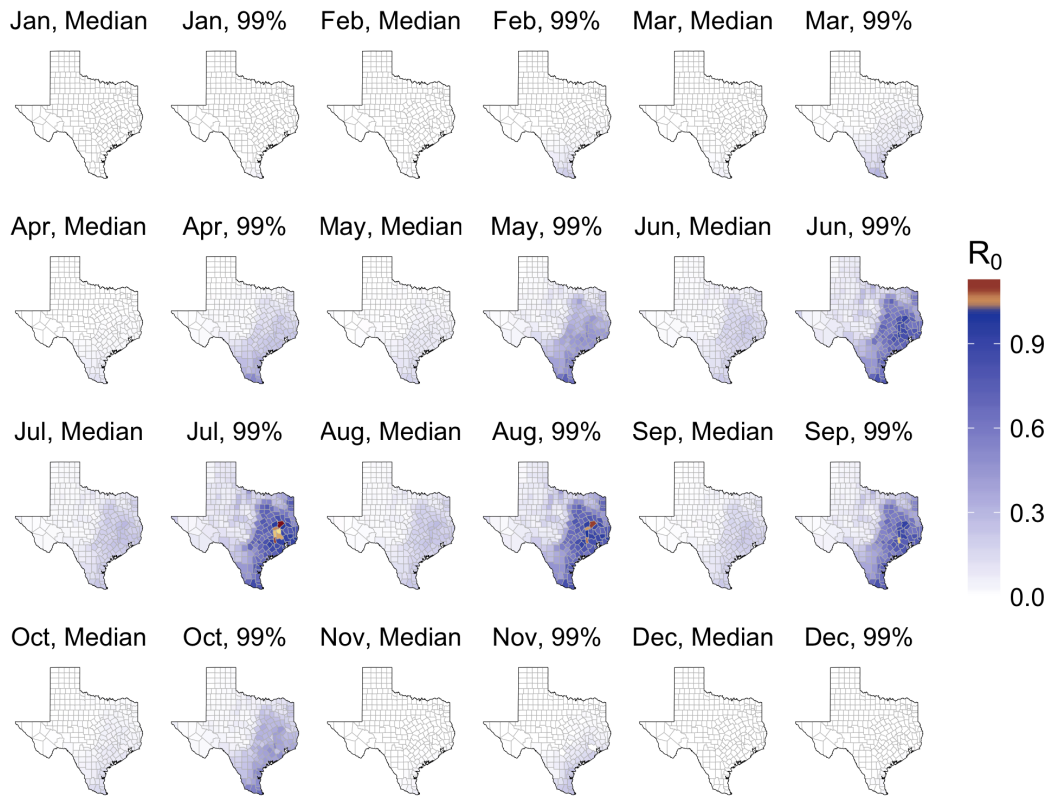
Figure 4.10: Posterior R_0 estimates for Texas counties for each month using the actual 2016 temperatures. Fill color indicates the estimated Median or 99 percentile estimate for that county for the given month, with counties showing yellow or red indicating their R_0 is above one (labels). Estimates are made using all importations through December of 2016, and assuming a single transmission event in both November and December.
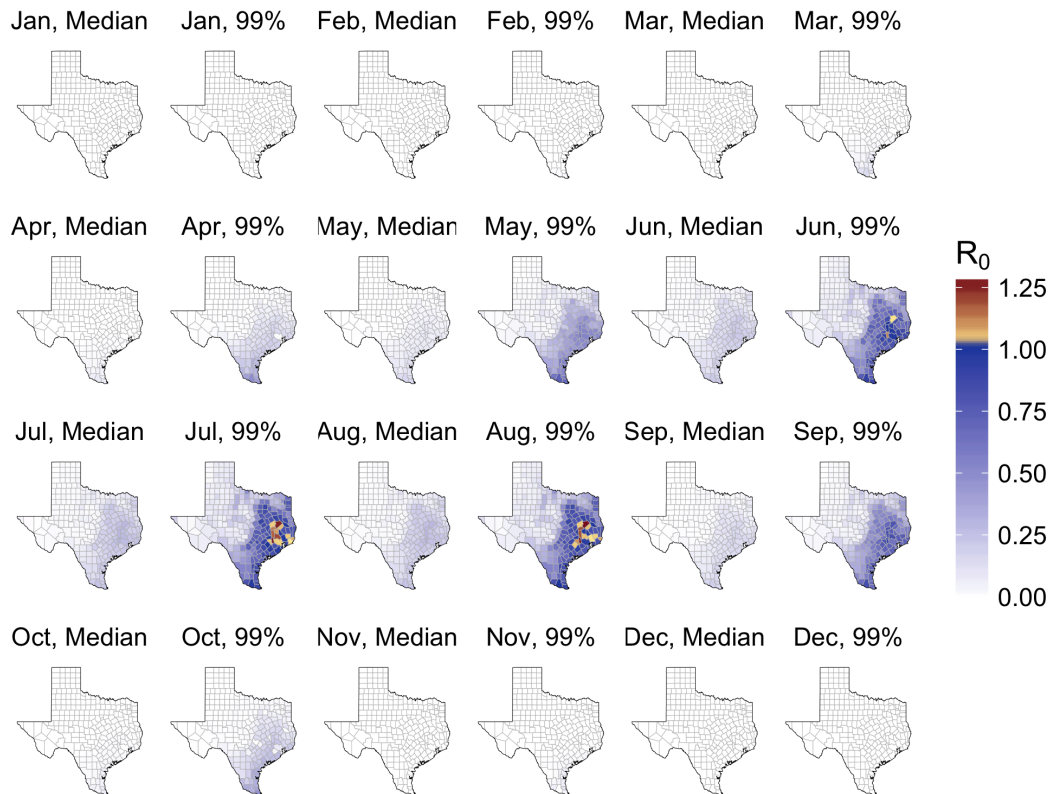
Figure 4.11: Posterior R_0 estimates for Texas counties for each month using historic temperature data for running estimation procedure and priors. Fill color indicates the estimated Median or 99 percentile estimate for that county for the given month, with counties showing yellow or red indicating their R_0 is above one (labels). Estimates are made using all importations through December of 2016, and assuming a single transmission event in November.
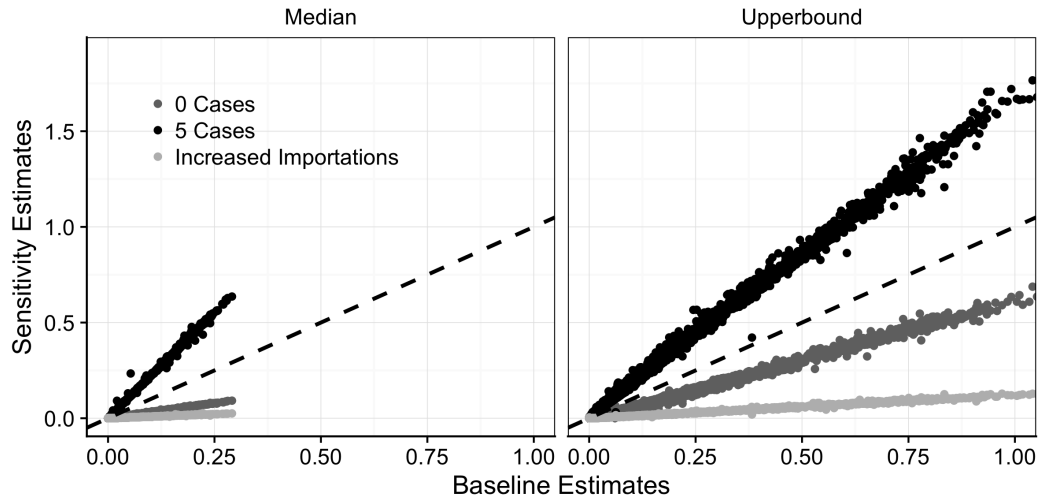
Figure 4.12: Sensitivity analysis of posterior R_0 estimates for each county. Each point indicates a county-month posterior R_0 estimate under different estimation scenarios. The x-axis value for all points is determined by the baseline scenario where posterior R_0 estimates consider only a single case detected in the November. The y-axis value is based on three sensitivity scenarios: (1) posterior estimates assuming no secondary transmission (dark grey), (2) five cases of secondary transmission in November (black), or (3) a single case of secondary transmission, but increased overall importations (light grey). Points falling above the black dashed line indicate that that a given scenario increases posterior R_0 estimates compared to baseline estimates, and points below the line indicate the opposite. Estimates are compared for the median (left), and the 99th percentile (right) of the county-month distributions. Posterior R_0 estimates increase if more secondary transmission is assumed, and decrease if less secondary transmission occurs, or is the absolute number of importations is increased.

Bibliography

[1] Estimated range of Aedes aegypti and Aedes albopictus in the United States, 2016.

[2] Hong Kong Battling Influenza Epidemic. *New York Times*, page 3, Apr 1957.

[3] Bulletin of the World Health Organization. Technical report, 1959.

[4] Natural history of dengue virus (DENV)-1 and DENV-4 infections: re-analysis of classic studies. *J Infect Dis [Internet]*, 195(7):1007–13, Apr [cited 2016 Apr 19] 2007.

[5] Texas Weather Data (http://www.usa.com/texas-state.htm), 2016.

[6] North america land data assimilation system (nldas) daily air temperatures and heat index, years 1979-2011, released 2012.

[7] T. Alex Perkins, Amir S. Siraj, Corrine W. Ruktanonchai, Moritz U. G. Kraemer, Andrew J. Tatem, J. Mlakar, P. Brasil, J. P. Messina, S. Cauchemez, M. A. Johansson, M. U. G. Kraemer, D. R. Lucey, L. O. Gostin, J. Elith, J. Leathwick, W. Kermack, A. McKendrick, A. Sorichetta, K. Sergon, D. L. Smith, O. J. Brady, M. Chan, M. A. Johansson, S. Funk, P. Reiter, V. Andreasen, K. A. Liebman, S. Bhatt, L. Feldstein, J. Brownstein, O. J. Brady, S. I. Hay, M. A. Johansson, A. J. Tatem, R. Hijmans,

W. Nordhaus, B. P. Taylor, C. J. E. Metcalf, A. J. Tatem, F. R. Stevens, D. Weiss, P. Gerardin, D. Wright, N. Pya, S. Wood, O. J. Brady, L. Muir, B. Kay, H. Nishiura, S. Halstead, T. W. Scott, R. C. Reiner, J. Lessler, J. Ma, D. J. Earn, M. U. G. Kraemer, M. Duffy, N. Schwarz, D. Sissoko, B. Dwibedi, N. Gay, M. Moro, and A. Balmaseda. Model-based projections of Zika virus infections in childbearing women in the Americas. *Nature Microbiology*, 1(9):16126, jul 2016.

[8] Jorge A Alfaro-Murillo, Alyssa S Parpia, Meagan C Fitzpatrick, Jules A Tamagnan, Jan Medlock, Martial L Ndeffo-Mbah, Durland Fish, María L Ávila-Agüero, Rodrigo Marín, Albert I Ko, and Alison P Galvani. A Cost-Effectiveness Tool for Informing Policies on Zika Virus Control. *PLOS Negl*, 10(5):e0004743, may 2016.

[9] Antoine Allard, Benjamin M Althouse, Laurent Hébert-Dufresne, and Samuel V Scarpino. The risk of sustained sexual transmission of zika is underestimated. *PLoS Pathog.*, 13(9):e1006633, September 2017.

[10] Wladimir J. Alonso, Cécile Viboud, Lone Simonsen, Eduardo W. Hirano, Luciane Z. Daufenbach, and Mark A. Miller. Seasonality of Influenza in Brazil: A Traveling Wave from the Amazon to the Subtropics. *American Journal of Epidemiology*, 165(12):1434–1442, jun 2007.

[11] Viggo Andreasen, Cécile Viboud, and Lone Simonsen. Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copen-

hagen: implications for pandemic control strategies. *The Journal of infectious diseases*, 197(2):270–278, 2008.

[12] Jason Asher, Christopher Barker, Grace Chen, Derek Cummings, Matteo Chinazzi, Shelby Daniel-Wayman, Marc Fischer, Neil Ferguson, Dean Follman, M Elizabeth Halloran, Michael Johansson, Kiersten Kugeler, Jennifer Kwan, Justin Lessler, Ira M Longini, Stefano Merler, Andrew Monaghan, Ana Pastore y Piontti, Alex Perkins, D Rebecca Prevots, Robert Reiner, Luca Rossi, Isabel Rodriguez-Barraquer, Amir S Siraj, Kaiyuan Sun, Alessandro Vespignani, and Qian Zhang. Preliminary results of models to predict areas in the americas with increased likelihood of zika virus transmission in 2017, 2017.

[13] Shweta Bansal, Bryan T Grenfell, and Lauren Ancel Meyers. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of The Royal Society Interface*, 4(16):879–891, oct 2007.

[14] Shweta Bansal and Lauren Ancel Meyers. The impact of past epidemics on future disease dynamics. *Journal of Theoretical Biology*, 309:176–84, sep 2012.

[15] Trevor Bedford, Andrew Rambaut, and Mercedes Pascual. Canalization of the evolutionary trajectory of the human influenza virus. *BMC biology*, 10(1):38, jan 2012.

[16] Trevor Bedford, Steven Riley, Ian G. Barr, Shobha Broor, Mandeep Chadha, Nancy J. Cox, Rodney S. Daniels, C Palani Gunasekaran, Aeron C Hurt, Anne Kelso, Alexander Klimov, Nicola S Lewis, Xiyan Li, John W. McCauley, Takato Odagiri, Varsha Potdar, Andrew Rambaut, Yuelong Shu, Eugene Skepner, Derek J. Smith, Marc A. Suchard, Masato Tashiro, Dayan Wang, Xiyan Xu, Philippe Lemey, and Colin A. Russell. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, 523(7559):217–220, jun 2015.

[17] W. I. B. Beveridge. Where did the red flu come from? *New Scientist*, pages 790–791, mar 1978.

[18] Matthew Biggerstaff, Simon Cauchemez, Carrie Reed, Manoj Gambhir, and Lyn Finelli. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC infectious diseases*, 14:480, 2014.

[19] William C Black, Kristine E Bennett, Norma Gorrochótegui-Escalante, Carolina V Barillas-Mury, Ildefonso Fernández-Salas, Marıa de Lourdes Muñoz, José A Farfán-Alé, Ken E Olson, and Barry J Beaty. Flavivirus susceptibility in aedes aegypti. *Archives of medical research*, 33(4):379–388, 2002.

[20] S. Blumberg and J.O. Lloyd-Smith. Comparing methods for estimating R0 from the size distribution of subcritical transmission chains. *Epidemics*, 5(3):131–145, sep 2013.

[21] Seth Blumberg and James O Lloyd-Smith. Inference of r(0) and transmission heterogeneity from the size distribution of stuttering chains. *PLoS Comput. Biol.*, 9(5):e1002993, May 2013.

[22] Oliver J. Brady, Michael A. Johansson, Carlos A. Guerra, Samir Bhatt, Nick Golding, David M. Pigott, Hélène Delatte, Marta G. Grech, Paul T. Leisnham, Rafael Maciel-de Freitas, Linda M. Styer, David L. Smith, Thomas W. Scott, Peter W. Gething, and Simon I. Hay. Modelling adult aedes aegypti and aedes albopictus survival at different temperatures in laboratory and field settings. *Parasites & Vectors*, 6(1):1–12, 2013.

[23] COF Buckee, Katia Koelle, and Sunetra Gupta. The effects of host contact network structure on pathogen diversity and strain structure. *Proceedings of the ...*, 101(29):10839–10844, 2004.

[24] Colin J Carlson, Eric Dougherty, Mike Boots, Wayne Getz, and Sadie J Ryan. Consensus and conflict among ecological forecasts of zika virus outbreaks in the united states. *Sci. Rep.*, 8(1):4921, March 2018.

[25] Colin J Carlson, Eric R Dougherty, and Wayne Getz. An ecological assessment of the pandemic threat of zika virus. *PLoS Negl. Trop. Dis.*, 10(8):e0004968, August 2016.

[26] Keith Carlson. *Mathematical Modeling of Infectious Diseases with Latency: Homogeneous Mixing and Contact Network*. Masters of science, University of Central Florida, 2016.

[27] Lauren A Castro, Xi Chen, B. Dimitrov, Nedialko, and Lauren Ancel Meyers. Texas arbovirus risk.

[28] Lauren A Castro, Spencer J Fox, Xi Chen, Kai Liu, Steven E Bellan, Nedialko B Dimitrov, Alison P Galvani, and Lauren Ancel Meyers. Assessing real-time zika risk in the united states. *BMC Infect. Dis.*, 17(1):284, May 2017.

[29] Simon Cauchemez, Scott Epperson, Matthew Biggerstaff, David Swerdlow, Lyn Finelli, and Neil M. Ferguson. Using Routine Surveillance Data to Estimate the Epidemic Potential of Emerging Zoonoses: Application to the Emergence of US Swine Origin Influenza A H3N2v Virus. *PLoS Medicine*, 10(3):e1001399, mar 2013.

[30] CDC. Update: Novel Influenza A (H1N1) Virus Infection — Mexico, March–May, 2009. Technical report, 2009.

[31] CDC. The 2009 H1N1 Pandemic: Summary Highlights, April 2009-April 2010. Technical report, CDC, 2010.

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

[33] Miranda Chan and Michael A Johansson. The incubation periods of dengue viruses. *PloS one*, 7(11):e50972, 2012.

[34] W K Chang. National influenza experience in Hong Kong, 1968. *Bull World Health Organ*, 41:349–351, 1969.

[35] Thais Chouin-Carneiro, Anubis Vega-Rua, Marie Vazeille, André Yebakima, Romain Girod, Daniella Goindin, Myrielle Dupont-Rouzeyrol, Ricardo Lourenço-de Oliveira, and Anna-Bella Failloux. Differential susceptibilities of aedes aegypti and aedes albopictus from the americas to zika virus. *PLoS Negl Trop Dis*, 10(3):e0004543, 2016.

[36] Gerardo Chowell, Luís M A Bettencourt, Niall Johnson, Wladimir J Alonso, and Cécile Viboud. The 1918-1919 influenza pandemic in England and Wales: spatial patterns in transmissibility and mortality impact. *Proceedings. Biological sciences / The Royal Society*, 275(1634):501–9, 2008.

[37] Gerardo Chowell, Santiago Echevarría-Zuno, Cecile Viboud, Lone Simonsen, James Tamerius, Mark A Miller, and Víctor H Borja-Aburto. Characterizing the Epidemiology of the 2009 Influenza A/H1N1 Pandemic in Mexico. *PLoS Medicine*, 8(5), 2011.

[38] Gerardo Chowell, Cécile Viboud, Lone Simonsen, Mark A Miller, and Rodolfo Acuna-Soto. Mortality patterns associated with the 1918 influenza pandemic in Mexico: evidence for a spring herald wave and lack of preexisting immunity in older populations. *The Journal of infectious diseases*, 202(4):567–75, 2010.

[39] Thomas S Churcher, Justin M Cohen, Joseph Novotny, Nyasatu Ntshal-intshali, Simon Kunene, and Simon Cauchemez. Public health. measuring the path toward malaria elimination. *Science*, 344(6189):1230–1232, June 2014.

[40] E C Claas, a D Osterhaus, R van Beek, J C De Jong, G F Rimmelzwaan, D a Senne, S Krauss, K F Shortridge, and R G Webster. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet (London, England)*, 351(9101):472–7, 1998.

[41] Nakia S Clemmons, Gregory S Wallace, Manisha Patel, and Paul A Gastañaduy. Incidence of measles in the united states, 2001-2015. *JAMA*, 318(13):1279, 2017.

[42] W C Cockburn, P J Delon, and W Ferreira. Origin and progress of the 1968-69 Hong Kong influenza epidemic. *Bull World Health Organ*, 41(3):345–8, 1969.

[43] Nicolas Collin and Xavier de Radiguès. Vaccine production capacity for seasonal and pandemic (H1N1) 2009 influenza. *Vaccine*, 27(38):5184–5186, aug 2009.

[44] Felipe J Colón-González, Carlos A Peres, Christine Steiner São Bernardo, Paul R Hunter, and Iain R Lake. After the epidemic: Zika virus projections for latin america and the caribbean. *PLoS Negl. Trop. Dis.*, 11(11):e0006007, November 2017.

[45] Benjamin J. Cowling, Lianmei Jin, Eric HY Lau, Qiaohong Liao, Peng Wu, Hui Jiang, Tim K. Tsang, Jiandong Zheng, Vicky J. Fang, Zhaorui Chang, Michael Y. Ni, Qian Zhang, Dennis KM Ip, Jianxing Yu, Yu Li, Liping Wang, Wenxiao Tu, Ling Meng, Joseph T. Wu, Huiming Luo, Qun Li, Yuelong Shu, Zhongjie Li, Zijian Feng, Weizhong Yang, Yu Wang, Gabriel M. Leung, and Hongjie Yu. Comparative epidemiology of human infections with avian influenza A H7N9 and H5N1 viruses in China: a population-based study of laboratory-confirmed cases. *The Lancet*, 382(9887):129–137, jul 2013.

[46] EJ Dechant and JG Rigau-Perez. Hospitalizations for suspected dengue in Puerto Rico, 1991-1995: estimation by capture-recapture methods. The Puerto Rico Association of Epidemiologists. *Am J Trop Med Hyg*, 61(4):574–578, oct 1999.

[47] Linh Dinh, Gerardo Chowell, Kenji Mizumoto, and Hiroshi Nishiura. Estimating the subcritical transmissibility of the zika outbreak in the state of florida, USA, 2016. *Theor. Biol. Med. Model.*, 13(1):20, November 2016.

[48] Ilaria Dorigatti, Simon Cauchemez, and Neil M Ferguson. Increased transmissibility explains the third wave of infection by the 2009 H1N1 pandemic virus in England. *Proceedings of the National Academy of Sciences of the United States of America*, 110(33):13422–7, aug 2013.

[49] Eve Dubé, Maryline Vivion, and Noni E. MacDonald. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: Influence, impact and implications. *Expert Review of Vaccines*, 14(1):99–117, 2014.

[50] Mark R. Duffy, Tai-Ho Chen, W. Thane Hancock, Ann M. Powers, Jacob L. Kool, Robert S. Lanciotti, Moses Pretrick, Maria Marfel, Stacey Holzbauer, Christine Dubray, Laurent Guillaumot, Anne Griggs, Martin Bel, Amy J. Lambert, Janeen Laven, Olga Kosoy, Amanda Panella, Brad J. Biggerstaff, Marc Fischer, and Edward B. Hayes. Zika virus outbreak on Yap Island, Federated States of Micronesia. *New Eng J Med [Internet]*, 360(24):2536–2543, Jun [cited 2016 Apr 4] 2009.

[51] Lizette O. Durand, Patrick Glew, Diane Gross, Matthew Kasper, Susan Trock, Inkyu K. Kim, Joseph S. Bresee, Ruben Donis, Timothy M. Uyeki, Marc-Alain Widdowson, and Eduardo Azziz-Baumgartner. Timing of Influenza A(H5N1) in Poultry and Humans and Seasonal Influenza Activity Worldwide, 2004-2013. *Emerging Infectious Diseases*, 21(2):202–208, feb 2015.

[52] Ali H Ellebedy, Florian Krammer, Gui-Mei Li, Matthew S Miller, Christopher Chiu, Jens Wrammert, Cathy Y Chang, Carl W Davis, Megan McCausland, Rivka Elbein, Srilatha Edupuganti, Paul Spearman, Sarah F Andrews, Patrick C Wilson, Adolfo García-Sastre, Mark J Mulligan, Aneesh K Mehta, Peter Palese, and Rafi Ahmed. Induction of broadly cross-reactive antibody responses to the influenza HA stem region follow-

ing H5N1 vaccination in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 111(36):13133–8, 2014.

[53] Luis E Escobar and Meggan E Craft. Advances and limitations of disease biogeography using ecological niche modeling. *Front. Microbiol.*, 7:1174, August 2016.

[54] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltán Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, may 2004.

[55] Michelle V Evans, Tad A Dallas, Barbara A Han, Courtney C Murdock, and John M Drake. Data-driven identification of potential zika virus vectors. *eLife*, 6:e22053, feb 2017.

[56] FAO-OIE-WHO. GLEWS+ The Joint FAO-OIE-WHO Global Early Warning System for health threats and emerging risks at the human-animal-ecosystems interface. 2013.

[57] Neil M Ferguson, Alison P. Galvani, and Robin M Bush. Ecological and immunological determinants of influenza evolution. *Nature*, 422(6930):428–433, mar 2003.

[58] Rosilainy Surubi Fernandes, Stéphanie Silva Campos, Anielly Ferreira-de Brito, Rafaella Moraes de Miranda, Keli Antunes Barbosa da Silva,

Marcia Gonçalves de Castro, Lidiane MS Raphael, Patrícia Brasil, Anna-Bella Failloux, Myrna C Bonaldo, et al. Culex quinquefasciatus from rio de janeiro is not competent to transmit the local zika virus. *PLoS Negl Trop Dis*, 10(9):e0004993, 2016.

[59] Matthew J Ferrari, Shweta Bansal, Lauren A Meyers, and Ottar N Bjørnstad. Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1602):2743–2748, 2006.

[60] Brian S Finkelman, Cecile Viboud, Katia Koelle, Matthew J Ferrari, Nita Bharti, and Bryan T Grenfell. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PloS one*, 2(12):e1296, jan 2007.

[61] Florida Department of Health. Department of Health Daily Zika Update. Technical report.

[62] H Fukumi. Summary report on the Asian influenza epidemic in Japan, 1957. *Bulletin of the World Health Organization*, 20(2-3):187–98, 1959.

[63] Catherine Gerdil. The annual production cycle for influenza vaccine. *Vaccine*, 21(16):1776–1779, may 2003.

[64] Daniel T Gillesple. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[65] Julia R Gog, Sébastien Ballesteros, Cecile Viboud, Lone Simonsen, Ottar N Bjørnstad, Jeffrey Shaman, Dennis L Chao, Farid Khan, and Bryan T Grenfell. Spatial Transmission of 2009 Pandemic Influenza in the US. *PLoS computational biology*, 10(6):e1003635, jun 2014.

[66] Edward Goldstein, Cecile Viboud, Vivek Charu, and Marc Lipsitch. Improving the estimation of influenza-related mortality over a seasonal baseline. *Epidemiology (Cambridge, Mass.)*, 23(6):829–38, nov 2012.

[67] Katelyn M Gostic, Monique R Ambrose, Michael Worobey, and James O Lloyd-Smith. Potent Protection against H5N1 and H7N9 Influenza via Childhood Hemagglutinin Imprinting. Technical report, jul 2016.

[68] Kristie M. Grebe, Jonathan W. Yewdell, and Jack R. Bennink. Heterosubtypic immunity to influenza A virus: where do we stand? *Microbes and Infection*, 10(9):1024–1029, 2008.

[69] Ardath Grills, Stephanie Morrison, Bradley Nelson, Jennifer Miniota, Alexander Watts, and Martin S Cetron. Projected zika virus importation and subsequent ongoing transmission after travel to the 2016 olympic and paralympic games — Country-Specific assessment, july 2016. *MMWR Morb. Mortal. Wkly. Rep.*, 65(28):711–715, 2016.

[70] Nathan D Grubaugh, Jason T Ladner, Moritz U G Kraemer, Gytis Dudas, Amanda L Tan, Karthik Gangavarapu, Michael R Wiley, Stephen White, Julien Thézé, Diogo M Magnani, Karla Prieto, Daniel Reyes,

Andrea M Bingham, Lauren M Paul, Refugio Robles-Sikisaka, Glenn Oliveira, Darryl Pronty, Carolyn M Barcellona, Hayden C Metsky, Mary Lynn Baniecki, Kayla G Barnes, Bridget Chak, Catherine A Freije, Adrianne Gladden-Young, Andreas Gnirke, Cynthia Luo, Bronwyn MacInnis, Christian B Matranga, Daniel J Park, James Qu, Stephen F Schaffner, Christopher Tomkins-Tinch, Kendra L West, Sarah M Winnicki, Shirlee Wohl, Nathan L Yozwiak, Joshua Quick, Joseph R Fauver, Kamran Khan, Shannon E Brent, Robert C Reiner, Jr, Paola N Lichtenberger, Michael J Ricciardi, Varian K Bailey, David I Watkins, Marshall R Cone, Edgar W Kopp, 4th, Kelly N Hogan, Andrew C Cannons, Reynald Jean, Andrew J Monaghan, Robert F Garry, Nicholas J Loman, Nuno R Faria, Mario C Porcelli, Chalmers Vasquez, Elyse R Nagle, Derek A T Cummings, Danielle Stanek, Andrew Rambaut, Mariano Sanchez-Lockhart, Pardis C Sabeti, Leah D Gillis, Scott F Michael, Trevor Bedford, Oliver G Pybus, Sharon Isern, Gustavo Palacios, and Kristian G Andersen. Genomic epidemiology reveals multiple introductions of zika virus into the united states. *Nature*, 546(7658):401–405, June 2017.

[71] Anne Gulland. Zika virus is a global public health emergency, declares WHO. *BMJ*, 352, 2016.

[72] Micah B Hahn, Rebecca J Eisen, Lars Eisen, Karen A Boegler, Chester G Moore, Janet McAllister, Harry M Savage, and John-Paul Mutebi. Reported distribution of aedes (stegomyia) aegypti and aedes (stegomyia)

albopictus in the united states, 1995-2016 (diptera: Culicidae). *J. Med. Entomol.*, June 2016.

[73] Rune Halvorsen, Sabrina Mazzoni, Anders Bryn, and Vegar Bakkestuen. Opportunities for improved distribution modelling practice via a strict maximum likelihood interpretation of maxent. *Ecography*, 38(2):172–183, 2015.

[74] Jennifer R Hamilton, David Sachs, Jean K Lim, Ryan A Langlois, Peter Palese, and Nicholas S Heaton. Club cells surviving influenza A virus infection induce temporary nonspecific antiviral immunity. *Proceedings of the National Academy of Sciences*, 113(14):3861–3866, apr 2016.

[75] J M Heffernan, R J Smith, and L M Wahl. Perspectives on the basic reproductive ratio. *J. R. Soc. Interface*, 2(4):281–293, September 2005.

[76] S Herfst, E J Schrauwen, M Linster, S Chutinimitkul, E de Wit, V J Munster, E M Sorrell, T M Bestebroer, D F Burke, D J Smith, G F Rimmelzwaan, A D Osterhaus, and R A Fouchier. Airborne transmission of influenza A/H5N1 virus between ferrets. *Science*, 336(6088):1534–1541, 2012.

[77] Pilar A Hernandez, Catherine H Graham, Lawrence L Master, and Deborah L Albert. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29(5):773–785, 2006.

[78] Robert J Hijmans, Susan E Cameron, Juan L Parra, Peter G Jones, Andy Jarvis, et al. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*, 25(15):1965–1978, 2005.

[79] G. K. Hirst. The Quantitative Determination of Influenza Virus and Antibodies By Means of Red Cell Agglutination. *Journal of Experimental Medicine*, 75(1):49–64, jan 1942.

[80] Thomas Hladish, Eugene Melamud, Luis Alberto Barrera, Alison P. Galvani, and Lauren Ancel Meyers. EpiFire: An open source C++ library and application for contact network epidemiology. *BMC bioinformatics*, 13:76, jan 2012.

[81] Homeland Security Council. Pandemic Influenza Implementation Plan. 2006.

[82] Masaki Imai, Tokiko Watanabe, Masato Hatta, Subash C. Das, Makoto Ozawa, Kyoko Shinya, Gongxun Zhong, Anthony Hanson, Hiroaki Katsura, Shinji Watanabe, Chengjun Li, Eiryo Kawakami, Shinya Yamada, Maki Kiso, Yasuo Suzuki, Eileen a. Maher, Gabriele Neumann, and Yoshihiro Kawaoka. Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature*, 486(7403):420–428, 2012.

[83] Charlotte Jackson, Emilia Vynnycky, Jeremy Hawker, Babatunde Olowokure, and Punam Mangtani. School closures and influenza: systematic review of epidemiological studies. *BMJ Open*, 3(2):e002149, feb 2013.

[84] E. T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, may 1957.

[85] Michael A. Jhung, David Swerdlow, Sonja J. Olsen, Daniel Jernigan, Matthew Biggerstaff, Laurie Kamimoto, Krista Kniss, Carrie Reed, Alicia Fry, Lynnette Brammer, Jacqueline Gindler, William J. Gregg, Joseph Bresee, and Lyn Finelli. Epidemiology of 2009 pandemic influenza a (H1N1) in the United States. *Clinical Infectious Diseases*, 52(SUPPL. 1):13–26, 2011.

[86] Niall P. A. S. Johnson and Juergen Mueller. Updating the Accounts: Global Mortality of the 1918-1920 "Spanish" Influenza Pandemic. *Bulletin of the History of Medicine*, 76(1):105–115, 2002.

[87] Edwin O. Jordan. *Epidemic Influenza: A Survey.* American Medical Association, Chicago, 1927.

[88] Jagat Narain Kapur and Hiremaglur K Kesavan. *Entropy optimization principles with applications.* Academic Pr, 1992.

[89] H. Kelly, S. Barry, K. Laurie, and G. Mercer. Seasonal influenza vaccination and the risk of infection with pandemic influenza: a possible illustration of non-specific temporary immunity following infection. *Euro surveillance: European communicable disease bulletin*, 15(47):1–6, nov 2010.

[90] Eben Kenah, Dennis L. Chao, Laura Matrajt, M. Elizabeth Halloran, and Ira M. Longini. The Global Transmission and Control of Influenza. *PLoS ONE*, 6(5):e19515, may 2011.

[91] Eben Kenah and James M. Robins. Second look at the spread of epidemics on networks. *Physical Review E*, 76(3):036113, 2007.

[92] Edwin D Kilbourne. Influenza pandemics of the 20th century. *Emerging infectious diseases*, 12(1):9–14, jan 2006.

[93] Moritz U G Kraemer, Marianne E Sinka, Kirsten A Duda, Adrian Mylne, Freya M Shearer, Christopher M Barker, Chester G Moore, Roberta G Carvalho, Giovanini E Coelho, Wim Van Bortel, Guy Hendrickx, Francis Schaffner, Iqbal RF Elyazar, Hwa-Jen Teng, Oliver J Brady, Jane P Messina, David M Pigott, Thomas W Scott, David L Smith, GR W Wint, Nick Golding, and Simon I Hay. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife [Internet]*, 4:e08347, Jun [cited 2016 Apr 19] 2015.

[94] A. Kucharski and J. R. Gog. Influenza emergence in the face of evolutionary constraints. *Proceedings of the Royal Society B: Biological Sciences*, 279(1729):645–652, 2012.

[95] Adam J Kucharski, Viggo Andreasen, and Julia R Gog. Capturing the dynamics of pathogens with many strains. *Journal of mathematical biology*, pages 1–24, 2015.

[96] Adam J Kucharski, Sebastian Funk, Rosalind M Eggo, Henri-Pierre Mallet, W John Edmunds, and Eric J Nilles. Transmission dynamics of zika virus in island populations: a modelling analysis of the 2013–14 french polynesia outbreak. *PLOS Negl Trop Dis*, 10(5):e0004726, 2016.

[97] Adam J Kucharski, Harriet L Mills, Christl A Donnelly, and Steven Riley. Transmission Potential of Influenza A(H7N9) Virus, China, 20132014. *Emerging Infectious Diseases*, 21(5):852–855, may 2015.

[98] A.J. Kucharski and W. J. Edmunds. Cross-immunity and age patterns of influenza A(H5N1) infection. *Epidemiology and Infection*, 143(06):1119–1124, apr 2015.

[99] Linda C Lambert and Anthony S Fauci. Influenza Vaccines for the Future. *New England Journal of Medicine*, 363(21):2036–2044, nov 2010.

[100] Karen L. Laurie, Louise A. Carolan, Deborah Middleton, Sue Lowther, Anne Kelso, and Ian G. Barr. Multiple Infections with Seasonal Influenza A Virus Induce CrossProtective Immunity against A(H1N1) Pandemic Influenza Virus in a Ferret Model. *The Journal of Infectious Diseases*, 202(7):1011–1020, oct 2010.

[101] Karen L. Laurie, Teagan A. Guarnaccia, Louise A. Carolan, Ada W C Yan, Malet Aban, Stephen Petrie, Pengxing Cao, Jane M. Heffernan, Jodie McVernon, Jennifer Mosse, Anne Kelso, James M. McCaw, and

Ian G. Barr. Interval Between Infections and Viral Hierarchy Are Determinants of Viral Interference Following Influenza Virus Infection in a Ferret Model. *Journal of Infectious Diseases*, 212(11):1701–1710, dec 2015.

[102] Justin Lessler, Cassandra T Ott, Andrea C Carcelen, Jacob M Konikoff, Joe Williamson, Qifang Bi, Nicholas G Reich, Derek A T Cummings, Lauren M Kucirka, and Lelia H Chaisson. Times to Key Events in the Course of Zika Infection and their Implications for Surveillance: A Systematic Review and Pooled Analysis. *bioRxiv [Internet]*, Mar [cited 2016 Apr 26] 2016.

[103] Gabriel E Leventhal, Alison L Hill, Martin a Nowak, and Sebastian Bonhoeffer. Evolution and emergence of infectious diseases in theoretical and real-world networks. *Nature Communications*, 2:1–11, 2015.

[104] Gui-Mei Li, Christopher Chiu, Jens Wrammert, Megan McCausland, Sarah F Andrews, Nai-Ying Zheng, Jane-Hwei Lee, Min Huang, Xinyan Qu, Srilatha Edupuganti, Mark Mulligan, Suman R Das, Jonathan W Yewdell, Aneesh K Mehta, Patrick C Wilson, and Rafi Ahmed. Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23):9047–52, 2012.

[105] Qun Li, Lei Zhou, Minghao Zhou, Zhiping Chen, Furong Li, Huanyu Wu, Nijuan Xiang, Enfu Chen, Fenyang Tang, Dayan Wang, Ling Meng, Zhiheng Hong, Wenxiao Tu, Yang Cao, Leilei Li, Fan Ding, Bo Liu, Mei Wang, Rongheng Xie, Rongbao Gao, Xiaodan Li, Tian Bai, Shumei Zou, Jun He, Jiayu Hu, Yangting Xu, Chengliang Chai, Shiwen Wang, Yongjun Gao, Lianmei Jin, Yanping Zhang, Huiming Luo, Hongjie Yu, Jianfeng He, Qi Li, Xianjun Wang, Lidong Gao, Xinghuo Pang, Guohua Liu, Yansheng Yan, Hui Yuan, Yuelong Shu, Weizhong Yang, Yu Wang, Fan Wu, Timothy M Uyeki, and Zijian Feng. Epidemiology of human infections with avian influenza A(H7N9) virus in china. *N. Engl. J. Med.*, 370(6):520–532, February 2014.

[106] S Liang, K Mozdzanowska, G Palladino, and W Gerhard. Heterosubtypic immunity to influenza type A virus in mice. Effector mechanisms and their longevity. *Journal of immunology (Baltimore, Md. : 1950)*, 152:1653–1661, 1994.

[107] Bruno Lina. History of Influenza Pandemics. *Paleomicrobiology: Past Human Infections.*, pages 199–211, 2008.

[108] Marc Lipsitch, Wendy Barclay, Rahul Raman, Charles J Russell, Jessica A. Belser, Sarah Cobey, Peter M. Kasson, James O. Lloyd-Smith, Sebastian Maurer-Stroh, Steven Riley, Catherine AA Beauchemin, Trevor Bedford, Thomas C. Friedrich, Andreas Handel, Sander Herfst, Pablo R. Murcia, Benjamin Roche, Claus O. Wilke, and Colin A. Russell. Viral

factors in influenza pandemic risk assessment. *eLife*, 5(NOVEMBER2016):1–38, nov 2016.

[109] Marc Lipsitch and Alison P Galvani. Ethical alternatives to experiments with novel potential pandemic pathogens. *PLoS medicine*, 11(5):e1001646, may 2014.

[110] Alun L. Lloyd. Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics. *Theoretical Population Biology*, 60(1):59–71, 2001.

[111] Christopher T. Lloyd, Alessandro Sorichetta, and Andrew J. Tatem. High resolution global gridded data for use in population studies. *Scientific Data*, 4:170001, jan 2017.

[112] J O Lloyd-Smith, S J Schreiber, P E Kopp, and W M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, November 2005.

[113] James O. Lloyd-Smith, Sebastian Funk, Angela R. McLean, Steven Riley, and James L N Wood. Nine challenges in modelling the emergence of novel pathogens. *Epidemics*, 10:35–39, 2014.

[114] Maimuna S Majumder, Emily Cohn, Durland Fish, and John S Brownstein. Estimating a feasible serial interval range for Zika fever. *Bulletin of the World Health Organization [Internet]*, 2016 [cited 2016 May 10].

[115] Alexander Mastin, Pablo Alarcon, Dirk Pfeiffer, James Wood, Susanna Williamson, Ian Brown, and Barbara Wieland. Prevalence and risk factors for swine influenza virus infection in the English pig population. *PLoS Currents*, 3:RRN1209, feb 2011.

[116] Stefano Merler, Marco Ajelli, Andrea Pugliese, and Neil M. Ferguson. Determinants of the Spatiotemporal Dynamics of the 2009 H1N1 Pandemic in Europe: Implications for Real-Time Modelling. *PLoS Computational Biology*, 7(9):e1002205, sep 2011.

[117] Cory Merow, Matthew J. Smith, and John A. Silander. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36(10):1058–1069, oct 2013.

[118] Jane P Messina, Moritz Ug Kraemer, Oliver J Brady, David M Pigott, Freya M Shearer, Daniel J Weiss, Nick Golding, Corrine W Ruktanonchai, Peter W Gething, Emily Cohn, John S Brownstein, Kamran Khan, Andrew J Tatem, Thomas Jaenisch, Christopher Jl Murray, Fatima Marinho, Thomas W Scott, and Simon I Hay. Mapping global environmental suitability for zika virus. *Elife*, 5, April 2016.

[119] Lauren Ancel Meyers. Contact Network Epidemiology: Bond Percolation Applied to Infectious Disease Prediction and Control. *Bulletin of the American Mathematical Society*, 44(1):63–86, 2007.

[120] Lauren Ancel Meyers, Babak Pourbohloul, M E J Newman, Danuta M

Skowronski, and Robert C Brunham. Network theory and SARS: predicting outbreak diversity. *Journal of theoretical biology*, 232(1):71–81, jan 2005.

[121] Joel C. Miller. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Physical Review E*, 76(1):010101, 2007.

[122] Joel C Miller. A note on a paper by Erik Volz: SIR dynamics in random networks. *Journal of mathematical biology*, 62(3):349–58, mar 2011.

[123] Joel C. Miller. Epidemics on networks with large initial conditions or changing structure. *PLoS ONE*, 9(7):1–9, 2014.

[124] Joel C Miller, Anja C Slim, and Erik M Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of The Royal Society Interface*, 9(70):890–906, may 2012.

[125] Joel C Miller and Erik M Volz. Incorporating Disease and Population Structure into Models of SIR Disease in Contact Networks. *PLoS ONE*, 8(8):e69162, aug 2013.

[126] Matthew S. Miller, Thomas J. Gardner, Florian Krammer, Lauren C. Aguado, Domenico Tortorella, Christopher F. Basler, and Peter Palese. Neutralizing Antibodies Against Previously Encountered Influenza Virus Strains Increase over Time: A Longitudinal Analysis. *Science translational medicine*, 5(198):198ra107, 2013.

[127] David M. Morens and Anthony S. Fauci. Emerging Infectious Diseases: Threats to Human Health and Global Stability. *PLoS Pathogens*, 9(7):7–9, 2013.

[128] Lynda E Muir and Brian H Kay. Aedes aegypti survival and dispersal estimated by mark-release-recapture in northern australia. *The American Journal of Tropical Medicine and Hygiene*, 58(3):277–282, 1998.

[129] Didier Musso, Van Mai Cao-Lormeau, and Duane J Gubler. Zika virus: following the path of dengue and chikungunya? *Lancet (London, England)*, 386(9990):243–4, jul 2015.

[130] Katsuhisa Nakajima, Ulrich Desselberger, and Peter Palese. Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature*, 274(5669):334–339, jul 1978.

[131] Martha I Nelson, Yi Tan, Elodie Ghedin, David E Wentworth, Kirsten St George, Laurel Edelman, Eric T Beck, Jiang Fan, Tommy Tsan-Yuk Lam, Swati Kumar, David J Spiro, Lone Simonsen, Cecile Viboud, Edward C Holmes, Kelly J Henrickson, and James M Musser. Phylogeography of the spring and fall waves of the H1N1/09 pandemic influenza virus in the United States. *Journal of virology*, 85(2):828–34, 2011.

[132] M. Newman. Threshold Effects for Two Pathogens Spreading on a Network. *Physical Review Letters*, 95(10):108701, sep 2005.

[133] Huan H Nguyen, Zina Moldoveanu, Miroslav J Novak, Frederik W Van Ginkel, Elisabeth Ban, Hiroshi Kiyono, Jerry R Mcghee, and Jiri Mestecky. Heterosubtypic Immunity to Lethal Influenza A Virus Infection Is Associated with Virus-Specific CD8+ Cytotoxic T Lymphocyte Responses Induced in Mucosa-Associated Tissues. *Virology*, 254:50–60, 1999.

[134] Huan H Nguyen, M. Zemlin, Ivaylo I Ivanov, Judit Andrasi, Cosima Zemlin, Huong L Vu, Robert Schelonka, Harry W Schroeder, and Jiri Mestecky. Heterosubtypic Immunity to Influenza A Virus Infection Requires a Properly Diversified Antibody Repertoire. *Journal of Virology*, 81(17):9331–9338, sep 2007.

[135] H. Nishiura. Mathematical and Statistical Analyses of the Spread of Dengue. *Dengue Bulletin [Internet]*, 30:51–67, 2006 [cited 2016 Apr 19].

[136] NOAA. Gridded Climate Datasets, 2017.

[137] William D Nordhaus. Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10):3510–3517, 2006.

[138] Office of Travel & Tourism Industries. US Monthly Arrivals Trend Line, 2014.

[139] Donald R Olson, Lone Simonsen, Paul J Edelson, and Stephen S Morse. Epidemiological evidence of an early wave of the 1918 influenza pandemic

in New York City. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31):11059–63, 2005.

[140] Michael T. Osterholm. Preparing for the Next Pandemic. *New England Journal of Medicine*, 352(18):1839–1842, may 2005.

[141] Harish Padmanabha, David Durham, Fabio Correa, Maria Diuk-Wasser, and Alison Galvani. The interactive roles of aedes aegypti super-production and human density in dengue transmission. *PLoS Negl. Trop. Dis.*, 6(8):e1799, August 2012.

[142] PAHO and WHO. Zika-Epidemiological Report - Anguilla. (August 4):1–2, 2016.

[143] PAHO and WHO. Zika-Epidemiological Report - Mexico. Technical Report August 5, 2016.

[144] Henry Franklin Parsons. *Influenza Epidemic of 1889-90.* Eyre and Spottiswoode, London, 1891.

[145] Natalie Pica, Rong Hai, Florian Krammer, Taia T Wang, Jad Maamary, Dirk Eggink, Gene S Tan, Jens C Krause, Thomas Moran, Cheryl R Stein, David Banach, Jens Wrammert, Robert B Belshe, Adolfo García-Sastre, and Peter Palese. Hemagglutinin stalk antibodies elicited by the 2009 pandemic influenza virus as a mechanism for the extinction of seasonal H1N1 viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 109(7):2573–8, 2012.

131

[146] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[147] Robert C Reiner, T Alex Perkins, Christopher M Barker, Tianchan Niu, Luis Fernando Chaves, Alicia M Ellis, Dylan B George, Arnaud Le Menach, Juliet RC Pulliam, Donal Bisanzio, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *Journal of The Royal Society Interface*, 10(81):20120921, 2013.

[148] Paul Reiter, Sarah Lathrop, Michel Bunning, Brad Biggerstaff, Daniel Singer, Tejpratap Tiwari, Laura Baber, Manuel Amador, Jaime Thirion, Jack Hayes, et al. Texas lifestyle limits transmission of dengue virus. *Emerging infectious diseases*, 9(1):86–89, 2003.

[149] Timothy P. Robinson, G. R. William Wint, Giulia Conchedda, Thomas P. Van Boeckel, Valentina Ercoli, Elisa Palamara, Giuseppina Cinardi, Laura D'Aietti, Simon I. Hay, and Marius Gilbert. Mapping the Global Distribution of Livestock. *PLoS ONE*, 9(5):e96084, may 2014.

[150] P Rohani, D J Earn, B. Finkenstadt, and B T Grenfell. Population dynamic interference among childhood diseases. *Proceedings of the Royal Society B: Biological Sciences*, 265(1410):2033–2041, nov 1998.

[151] Diana Patricia Rojas, Natalie E Dean, Yang Yang, Eben Kenah, Juliana Quintero, Simon Tomasi, Erika Lorena Ramirez, Yendi Kelly, Carolina Castro, Gabriel Carrasquilla, M Elizabeth Halloran, and Ira M Longini.

The epidemiology and transmissibility of Zika virus in Girardot and San Andres island, Colombia, September 2015 to January 2016. *Eurosurveillance*, 21(28):30283, jul 2016.

[152] Michelle Rozo and Gigi Kwik Gronvall. The Reemergent 1977 H1N1 Strain and the Gain-of-Function Debate. *mBio*, 6(4):e01013–15, sep 2015.

[153] Colin A Russell, Peter M Kasson, Ruben O Donis, Steven Riley, John Dunbar, Andrew Rambaut, Jason Asher, Stephen Burke, C Todd Davis, Rebecca J Garten, Sandrasegaram Gnanakaran, Simon I Hay, Sander Herfst, Nicola S Lewis, James O Lloyd-Smith, Catherine A Macken, Sebastian Maurer-Stroh, Elizabeth Neuhaus, Colin R Parrish, Kim M Pepin, Samuel S Shepard, David L Smith, David L Suarez, Susan C Trock, Marc-Alain Widdowson, Dylan B George, Marc Lipsitch, and Jesse D Bloom. Improving pandemic influenza risk assessment. *eLife*, 3:1–12, 2014.

[154] Samuel V Scarpino, Atila Iamarino, Chad Wells, Dan Yamin, Martial Ndeffo-Mbah, Natasha S Wenzel, Spencer J Fox, Tolbert Nyenswah, Frederick L Altice, Alison P Galvani, Lauren Ancel Meyers, and Jeffrey P Townsend. Epidemiological and viral genomic sequence analysis of the 2014 ebola outbreak reveals clustered transmission. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 60(7):1079–82, apr 2015.

[155] Judith R. Schäffr, Yoshihiro Kawaoka, William J Bean, Jochen Süss, Dennis Senne, and Robert G Webster. Origin of the Pandemic 1957 H2 Influenza A Virus and the Persistence of Its Possible Progenitors in the Avian Reservoir. *Virology*, 194(2):781–788, jun 1993.

[156] C. Scholtissek. Source for influenza pandemics. *European Journal of Epidemiology*, 10(4):455–458, 1994.

[157] J L Schulman and E D Kilbourne. Induction of Partial Specific Heterotypic Immunity in Mice By a Single Infection With Influenza a Virus. *Journal of bacteriology*, 89(1):170–4, 1965.

[158] Thomas W Scott, Priyanie H Amerasinghe, Amy C Morrison, Leslie H Lorenz, Gary G Clark, Daniel Strickman, Pattamaporn Kittayapong, and John D Edman. Longitudinal studies of aedes aegypti (diptera: Culicidae) in thailand and puerto rico: blood feeding frequency. *Journal of medical entomology*, 37(1):89–101, 2000.

[159] Jeffrey Shaman, Edward Goldstein, and Marc Lipsitch. Absolute humidity and pandemic versus epidemic influenza. *American journal of epidemiology*, 173(2):127–35, jan 2011.

[160] Jeffrey Shaman, Virginia E Pitzer, Cecile Viboud, Bryan T Grenfell, and Marc Lipsitch. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS biology*, 8(2):e1000316, feb 2010.

[161] Yue-Long Shu, Li-Qun Fang, Sake J. de Vlas, Yan Gao, Jan Hendrik Richardus, and Wu-Chun Cao. Dual Seasonal Patterns for Influenza, China. *Emerging Infectious Diseases*, 16(4):725–726, apr 2010.

[162] Tom J Sidwa. Mosquito Surveillance/Control in Texas. Technical report, 2016.

[163] Amir S Siraj and T Alex Perkins. Assessing the population at risk of zika virus in asia – is the emergency really over? *BMJ Global Health*, 2(3):e000309, 2017.

[164] Gavin J. D. Smith, Justin Bahl, Dhanasekaran Vijaykrishna, Jinxia Zhang, Leo L. M. Poon, Honglin Chen, Robert G. Webster, J. S. Malik Peiris, and Yi Guan. Dating the emergence of pandemic influenza viruses. *Proceedings of the National Academy of Sciences*, 106(28):11709–11712, jul 2009.

[165] Gavin J D Smith, Dhanasekaran Vijaykrishna, Justin Bahl, Samantha J Lycett, Michael Worobey, Oliver G Pybus, Siu Kit Ma, Chung Lam Cheung, Jayna Raghwani, Samir Bhatt, J S Malik Peiris, Yi Guan, and Andrew Rambaut. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*, 459(7250):1122–5, 2009.

[166] Karline Soetart, Thomas Petzoldt, and R. Woodrow Setzer. Solving Differential Equations in R: Package deSolve. *Journal of Statistical Software*, 33(9):1–25, 2010.

[167] Tadao Sonoguchi, H. Naito, Masaru Hara, Yasue Takeuchi, and Hideo Fukumi. Cross-subtype protection in humans during sequential, overlapping, and/or concurrent epidemics caused by h3n2 and h1n1 influenza viruses. *Journal of Infectious Diseases*, 151(1):81–88, 1985.

[168] G A Soper. The Influenza Pneumonia Pandemic in the American Army Camps during September and October. 1918. *Science*, 48(1245):451–456, 1918.

[169] K Subbarao, a Klimov, J Katz, H Regnery, W Lim, H Hall, M Perdue, D Swayne, C Bender, J Huang, M Hemphill, T Rowe, M Shaw, X Xu, K Fukuda, and N Cox. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science (New York, N.Y.)*, 279(5349):393–396, 1998.

[170] James Tamerius, Martha I. Nelson, Steven Z. Zhou, Cécile Viboud, Mark A. Miller, and Wladimir J. Alonso. Global Influenza Seasonality: Reconciling Patterns across Temperate and Tropical Regions. *Environmental Health Perspectives*, 119(4):439–445, nov 2010.

[171] James D. Tamerius, Jeffrey Shaman, Wladmir J. Alonso, Kimberly Bloom-Feshbach, Christopher K. Uejio, Andrew Comrie, and Cécile Viboud. Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. *PLoS Pathogens*, 9(3):e1003194, mar 2013.

[172] Jeffery K. Taubenberger and David M. Morens. 1918 Influenza: the Mother of All Pandemics. *Emerging Infectious Diseases*, 12(1):15–22, jan 2006.

[173] Koon K Teo, Janice Pogue, Leanne Dyal, Ingrid Copland, Population Health, Hamilton Health Sciences, Helmut Schumacher, Boehringer Ingelheim, Gilles Dagenais, United Kingdom, Craig Anderson, and Health Sciences. Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans. *New England Journal of Medicine*, 360(25):2605–2615, jun 2009.

[174] Texas Department of State Health and Human Services. Texas Announces Local Zika Virus Case in Rio Grande Valley. Technical report.

[175] Texas Department of State Health and Human Services. Arbovirus Activity in Texas 2013 Surveillance Report. Technical report, Texas Department of State Health Services; Infectious Disease Control Unit Zoonosis Control Branch, 2013.

[176] Texas Department of State Health and Human Services. Arbovirus Activity in Texas 2014 Surveillance Report. Technical report, Texas Department of State Health Services-Infectious Disease Control Unit Zoonosis Control Branch, 2014.

[177] Texas Department of State Health and Human Services. Zika in Texas [Internet], 2016 [cited 2016 Apr 19].

[178] Texas Department of State Health and Human Services. Zika in Texas, 2017.

[179] Damon J A Toth, Adi V Gundlapalli, Karim Khader, Warren B P Pettey, Michael A Rubin, Frederick R Adler, and Matthew H Samore. Estimates of outbreak risk from new introductions of ebola with immediate and delayed transmission control. *Emerg. Infect. Dis.*, 21(8):1402–1408, August 2015.

[180] A. R. Tuite, A. L. Greer, M. Whelan, A.-L. Winter, B. Lee, P. Yan, J. Wu, S. Moghadas, D. Buckeridge, B. Pourbohloul, and D. N. Fisman. Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. *Canadian Medical Association Journal*, 182(2):131–136, 2010.

[181] W. Tun-Lin, T. R. Burkot, and B. H. Kay. Effects of temperature and larval diet on development rates and survival of the dengue vector Aedes aegypti in north Queensland, Australia. *Med Vet Entomol [Internet]*, 14(1):31–37, Mar [cited 2016 Apr 19] 2000.

[182] Teja Turk, Nadine Bachmann, Claus Kadelka, Jürg Böni, Sabine Yerly, Vincent Aubert, Thomas Klimkait, Manuel Battegay, Enos Bernasconi, Alexandra Calmy, Matthias Cavassini, Hansjakob Furrer, Matthias Hoffmann, Huldrych F Günthard, Roger D Kouyos, Swiss HIV Cohort Study, V Aubert, M Battegay, E Bernasconi, J Böni, D L Braun, H C Bucher, A Calmy, M Cavassini, A Ciuffi, G Dollenmaier, M Egger, L Elzi,

J Fehr, J Fellay, H Furrer, C A Fux, H F Günthard, D Haerry, B Hasse, H H Hirsch, M Hoffmann, I Hösli, C Kahlert, L Kaiser, O Keiser, T Klimkait, R D Kouyos, H Kovari, B Ledergerber, G Martinetti, B Martinez de Tejada, C Marzolini, K J Metzner, N Müller, D Nicca, G Pantaleo, P Paioni, A Rauch, C Rudin, A U Scherrer, P Schmid, R Speck, M Stöckle, P Tarr, A Trkola, P Vernazza, G Wandeler, R Weber, and S Yerly. Assessing the danger of self-sustained HIV epidemics in heterosexuals by population based phylogenetic cluster analysis. *Elife*, 6, September 2017.

[183] U.S. Bureau of Economic Analysis. Texas Counties: Per Capita Income [Internet], 2014 [cited 2016 Apr 19].

[184] Cécile Viboud, Wladimir J Alonso, and Lone Simonsen. Influenza in Tropical Regions. *PLoS Medicine*, 3(4):e89, mar 2006.

[185] Erik Volz. SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology*, 56(3):293–310, 2008.

[186] J Wallinga and M Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, feb 2007.

[187] WHO. The top 10 causes of death.

[188] WHO. Global influenza surveillance and response system. 2013.

139

[189] WHO. 2018 Annual review of diseases prioritized under the Research and. *R&D Blueprint*, (February), 2018.

[190] WHO and CDC. National and Regional Level Outpatient Illness and Viral Surveillance, 2016.

[191] Paul S. Wikramaratna, Adam Kucharski, Sunetra Gupta, Viggo Andreasen, Angela R. McLean, and Julia R. Gog. Five challenges in modelling interacting strain dynamics. *Epidemics*, 10:31–34, 2015.

[192] Laurence A Wolsey. *Integer programming*, volume 42. Wiley New York, 1998.

[193] World Health Organization. Pandemic Influenza Preparedness and Response: A WHO Guidance Document. page 64, 2010.

[194] World Health Organization. WHO MERS-CoV global summary and assessment of risk. Technical report, July 2017.

[195] M. Worobey, G.-Z. Han, and A. Rambaut. Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proceedings of the National Academy of Sciences*, pages 1–6, apr 2014.

[196] David Hamilton Wright. Correlations Between Incidence and Abundance are Expected by Chance. *J Biogeogr [Internet]*, 18(4):463–466, 1991 [cited 2016 Apr 19].

[197] Wan Yang, Marc Lipsitch, and Jeffrey Shaman. Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences*, 3(11):201415012, 2015.

[198] Wan Yang, Elisaveta Petkova, and Jeffrey Shaman. The 1918 influenza pandemic in New York City: Age-specific timing, mortality, and transmission dynamics. *Influenza and other Respiratory Viruses*, 8(2):177–188, 2014.

[199] Qian Zhang, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore y Piontti, Natalie E Dean, Diana Patricia Rojas, Stefano Merler, Dina Mistry, Piero Poletti, Luca Rossi, Margaret Bray, M. Elizabeth Halloran, Ira M. Longini, and Alessandro Vespignani. Spread of Zika virus in the Americas. *Proceedings of the National Academy of Sciences*, 114(22):E4334–E4343, 2017.

[200] Daniel Zinder, Trevor Bedford, Sunetra Gupta, and Mercedes Pascual. The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. *PLoS pathogens*, 9(1):e1003104, jan 2013.

Vita

Spencer James Fox grew up in Lincolnwood, Illinois, where he attended Todd Hall, Rutledge Hall, and Lincoln Hall and honed his basketball skills on the slightly slanted courts of Henry A. Proesel park. He attended Niles West High School in Skokie and graduated in 2005. From there he went to Rose-Hulman Institute of Technology, and received a Bachelor of Science in Applied Biology in 2009. Following graduation, he chose to take his talents (basketball and otherwise) to the University of Texas at Austin, and enrolled in the Ecology, Evolution, and Behavior doctoral graduate program.

Permanent address: 6749 N Keeler
                   Lincolnwood, Illinois 60712

This dissertation was typed by Spencer James Fox