

# The unseen and pervasive threat of COVID-19 throughout the US

## Appendix

### Methods

#### *County data*

We obtained county-level estimates for COVID-19 cases from a data repository curated by the New York Times (1).

#### *Model*

We adapted the framework of ref. (2) to model COVID-19 in US counties. It assumes a branching process model for early transmission in which the number of secondary infections per infected case is distributed according to a negative binomial distribution to capture occasional superspreading events, as estimated for SARS (3). We account for imperfect detection and COVID-19 specific epidemiological characteristics (details in Table).

We run 100,000 stochastic outbreak simulations beginning with a single unreported case and ending when the cumulative cases reach 2000 or the outbreak dies out (whichever comes first). Following the methodology that we developed in ref. (2), outbreaks that reach 2000 cases and reach a minimum prevalence of 50 cases in a given day are classified as epidemics. We calculate the probability of sustained community transmission (required to cause an epidemic) for a given number of reported cases,  $x$ , by looking at all outbreaks that had  $x$  reported cases, and calculating the proportion of those outbreaks that progressed to epidemics (Figure S1). For outbreaks that became epidemics, we also calculated the distribution of lags (in weeks) between the day that the  $x$ th case is reported and the day that the epidemic surpasses 1,000 cumulative cases (Figure S2). We matched county case numbers with the reported case number to obtain epidemic probabilities for each of the US counties based on their current number of reported cases. Due to the intense travel restrictions in place in the US currently, we did not model imported cases between counties. Once mobility data becomes available the modeling framework has the ability to incorporate those estimates, which will likely raise epidemic risk across all counties.

#### *Sensitivity Analysis*

Our baseline scenario assumes the  $R_0$  of COVID-19 is 1.5 (accounting for ongoing social distancing measures across the US (4)) and that 10% of all cases are reported. To assess the impact of these assumptions on our estimates, we conducted a sensitivity analysis that varied  $R_0$  (1.1 and 3) and reporting rates (5%-40%). Generally for a given number of reported cases, higher transmission rates and lower reporting rates increase the estimated local risk of sustained transmission, while lower transmission rates and higher reporting rates reduce estimates (Figure S3).

Table. Model parameters used for simulating county COVID-19 outbreaks

Parameter	Description	Estimate	Source
$R_e$	Effective reproduction number: Average number of new cases from one infected individual in a susceptible population	1.5	(5)
$T_G$	Generation time (days): Average length of time between consecutive exposures $T_G = \frac{e}{\nu} + \left(\frac{1}{2}\right)\frac{n}{\delta} = T_E + \left(\frac{1}{2}\right)T_I$	6	(6,7)
$T_E$	Latent period (days)	1.25	Fit to $T_G$
$T_I$	Infectious period (days)	9.5	(6)
$e$	Number of exposed compartments in boxcar implementation	1	Fit to $T_G$
$n$	Number of infectious compartments in boxcar implementation	7	(6)
$\nu$	Incubation rate: Daily probability of progressing from one exposed compartment to the next	0.80	Fit to $T_G$
$\delta$	Recovery rate: Daily probability of progressing from one infectious compartment to the next	0.73	Fit to $T_I$
$\eta$	Daily detecting rate: The daily probability of an infectious individual being detected $\frac{0.1}{T_I}$	0.01	(8)
$d_t$	Total dispersion parameter of negative binomial distribution	0.16	(3)
	R code for number of new infectious individuals drawn daily: $rnbinom(n = 1, prob = \frac{d_t}{R_e + d_t}, size = \frac{d_t}{T_I})$		

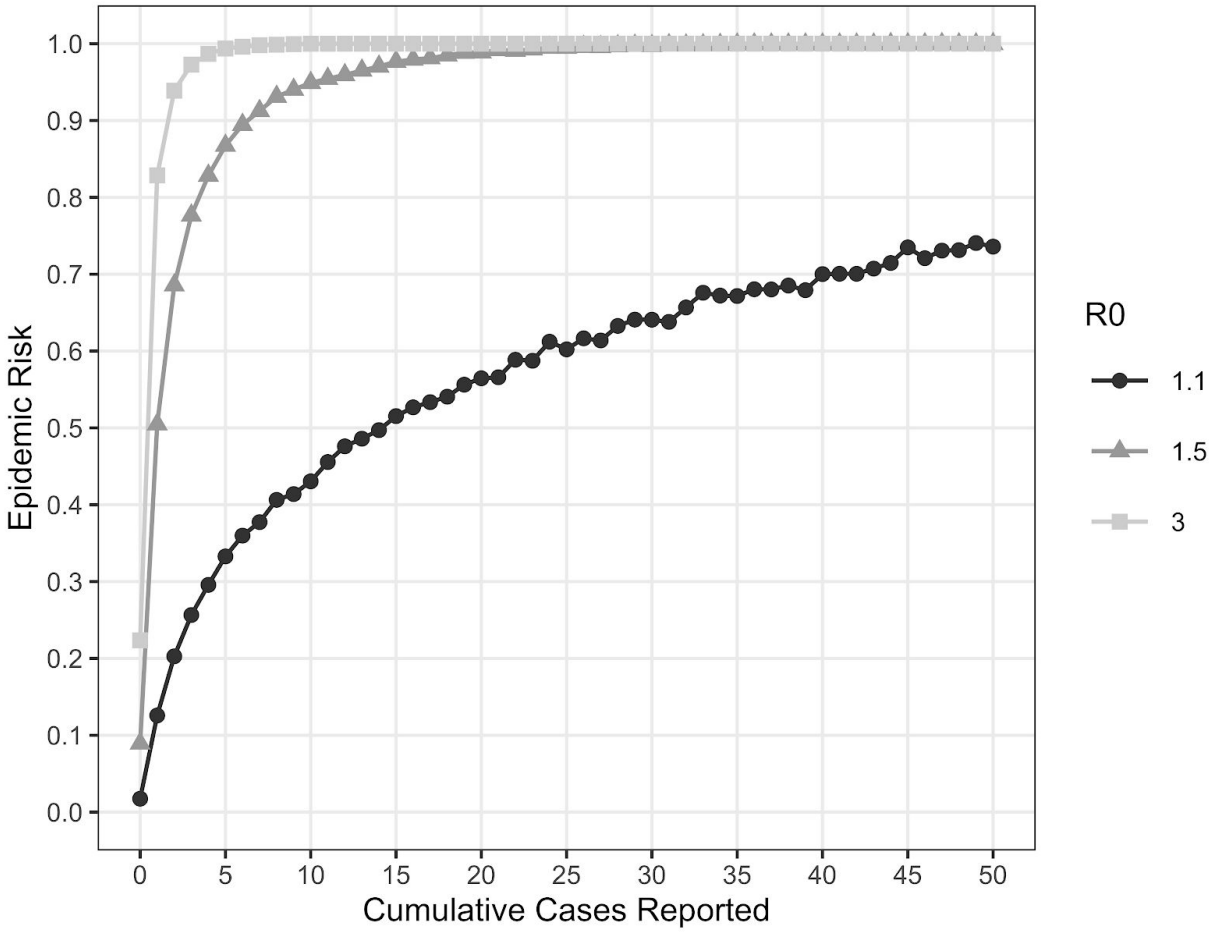


Figure S1. Sensitivity analysis with respect to the reproduction number ( $R_0$ ). For a given number of reported cases, the estimated risk of an epidemic increases with  $R_0$ . By the time a single case is reported, there is a 13%, 50% or 83% chance of an ongoing epidemic for  $R_0$  of 1.1, 1.5 or 3, respectively.

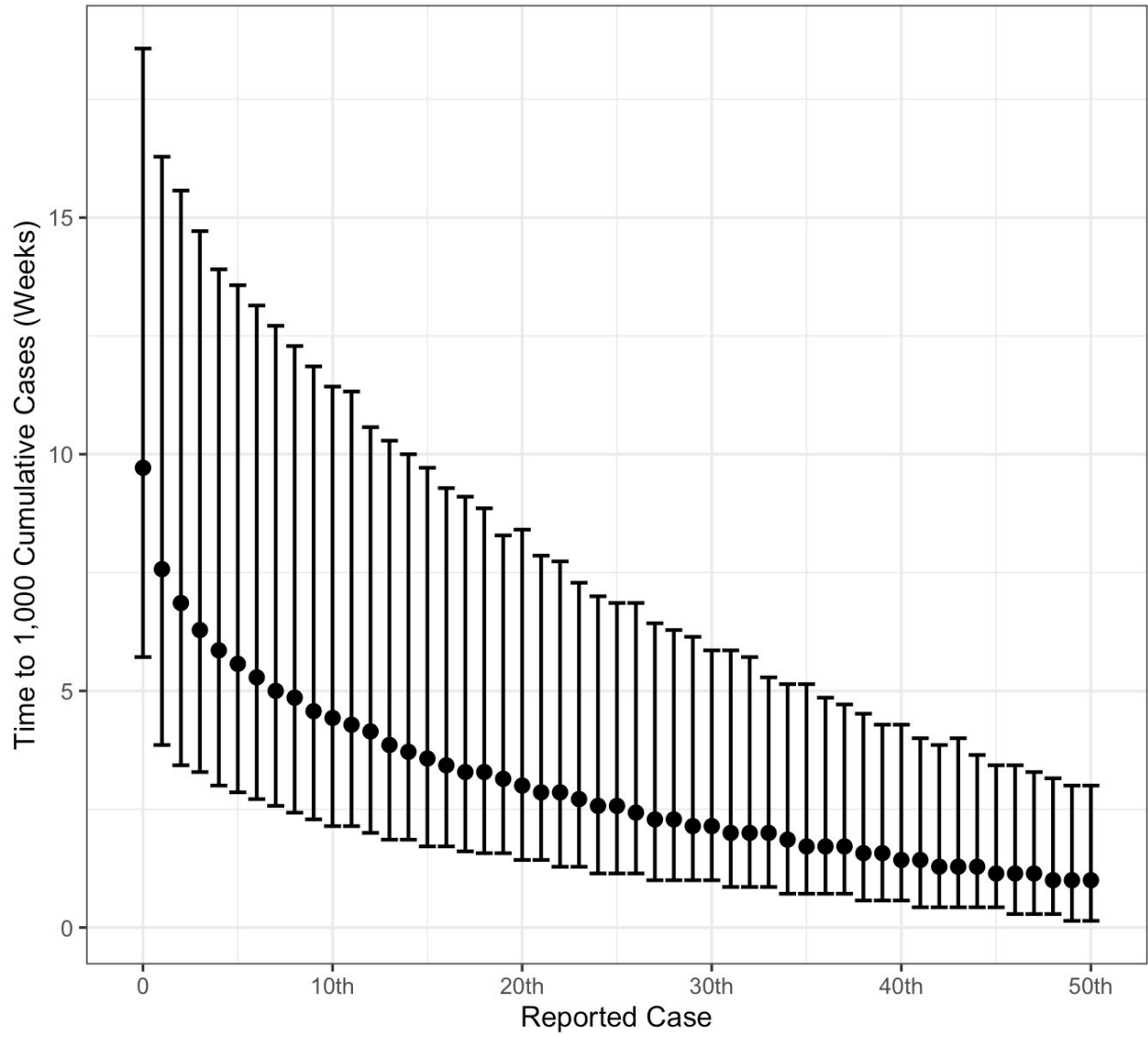


Figure S2. Time between the reporting of a case and the local epidemic exceeding 1,000 cumulative cases ( $R_0 = 1.5$ ). For a given number of cumulative reported cases (x-axis), we assume that an epidemic is underway and then estimate the median and 95% CI (error bars) number of weeks until the cumulative cases exceed 1,000. At the time the first case is reported, we would expect there to be 7.5 (95% CI 3.9-16.3) weeks before the cumulative cases surpass 1,000; from the date of the 10th reported case, this lead time shrinks to 4.4 (95% CI 2.1-11.4) weeks. These values are derived from 100,000 simulations.

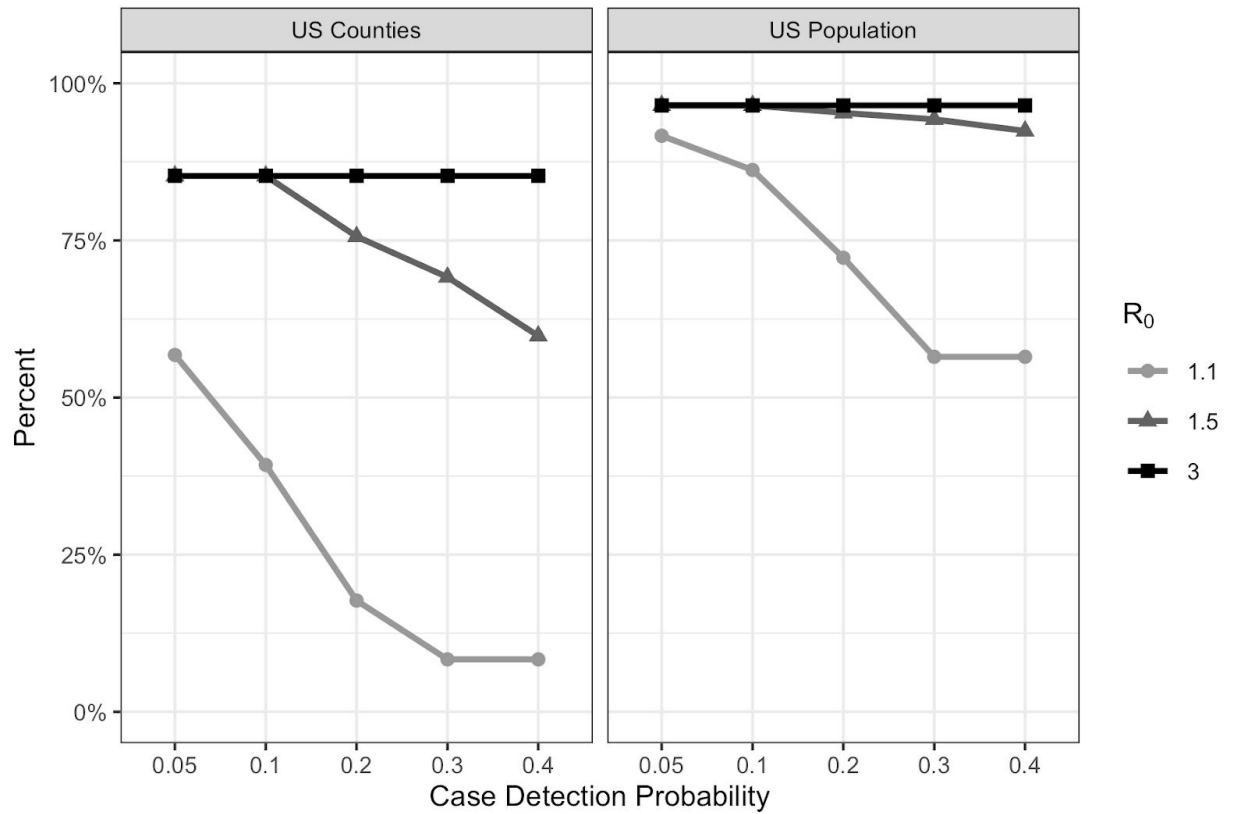


Figure S3. Sensitivity analysis with respect to the reproduction number ( $R_0$ ) and case detection probability. Percentage of US counties (left) or US population (right) that have greater than a 50% risk for sustained local transmission across varying assumed transmission rates (colors) and case detection probabilities (x-axis).

## References

1. Smith M, Yourish K, Almkhatar S, Collins K, Ivory D, Harmon A. The New York Times Coronavirus (Covid-19) Data in the United States. 2020 [cited 2020 Apr 16]. Available from: <https://github.com/nytimes/covid-19-data>
2. Castro LA, Fox SJ, Chen X, Liu K, Bellan SE, Dimitrov NB, et al. Assessing real-time Zika risk in the United States. *BMC Infect Dis*. 2017 May 4;17(1):284.
3. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005 Nov 17;438(7066):355–9.
4. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis*. 2020 Mar 23.
5. Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. *Int J Infect Dis*. 2020 Mar 17;93:339–44.
6. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. 2020.
7. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. 2020.
8. Perkins A, Cavany SM, Moore SM, Oidtman RJ, Lerch A, Poterek M. Estimating unobserved SARS-CoV-2 infections in the United States. 2020.

### **Code and Data Availability**

The original code adapted to this model is available as an R package at <https://github.com/sjfox/rtZIKVrisk>. Code and data for this paper are available at <https://github.com/MeyersLabUTexas/COVID-19-Epidemic-Risk>.