Yale SCHOOL *of* MANAGEMENT

YALE UNIVERSITY
School of Public Health

YALE UNIVERSITY
School of Engineering and
Applied Science

COVID-19 Scratch Models

Edward H. Kaplan

William N and Marie A Beach Professor of Operations Research

Yale School of Management

Professor of Public Health

Yale School of Public Health

Professor of Engineering

Yale School of Engineering and Applied Science

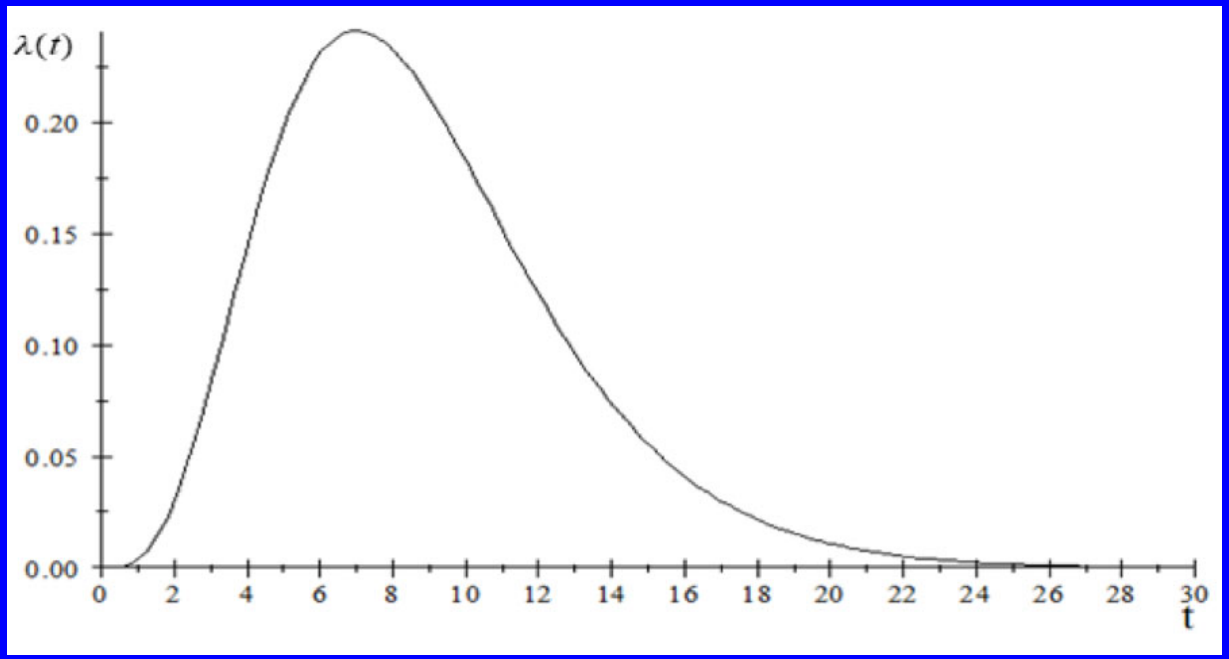
Columbia/Stanford OR PhD COVID-19 Seminar

Outline

- Basics: Euler-Lotka equation, R_0, scratch model, final size, herd immunity
- Aligning indicators via scratch model: basic convolution, application to hospital admissions and sludge RNA data (joint with Jordan Peccia, Yale School of Engineering)
- Interventions: testing and isolation, repeat testing, residential college models (joint with Joe Chang, Yale Data Science and Statistics, and Forrest Crawford, Yale Biostatistics)
- References on last slide

Early Transmission Basics

- Define $\lambda(u)$ as instantaneous transmission rate from a person already infected for duration u (think non-homogeneous Poisson)



$$R_0 = \int_0^\infty \lambda(u)du$$

R_0 is the *reproductive number* (expected number of infections transmitted per infected person early in outbreak)

Epidemic requires $R_0 > 1$

Early Outbreak Dynamics

- $\lambda(u)$ is transmission rate from person infected for duration u
- Once transmission starts, exponential growth in infections
- Incidence of *new* infections grows as ke^{rt} from i_0 initial infections

$$ke^{rt} = \int_0^t ke^{r(t-u)}\lambda(u)du + \iota_0\lambda(t), \quad t > 0$$

- Multiply both sides by e^{-rt}/k and let t grow yields ($e^{-rt}\lambda(t) \rightarrow 0$)

$$\int_0^\infty e^{-ru}\lambda(u)du = 1.$$

Early Outbreak Dynamics

- Euler Lotka equation $\int_0^\infty e^{-ru}\lambda(u)du = 1.$
- The integrand $b(u) = e^{-ru}\lambda(u)$ is a probability density – of all infections taking place now, what fraction are transmitted by people who were themselves infected u time units ago?
- $b(u)$ is the *backwards generation time density*; can be estimated from contact tracing data (look backwards from contacts to infectors)
- Exponential growth rate r estimated from initial growth in cases (.1/d)
- So write $\lambda(u) = e^{ru}b(u)$ and estimate $R_0 = \int_0^\infty \lambda(u)du$ = 2 to 3

Early Outbreak Dynamics

- If instead of working backwards, one starts with an *index case* and looking forward asks for the probability distribution of the time until said index infects a contact, you get the *forward generation time* density $f(u) = \lambda(u) / R_0$
- The Euler Lotka equation $\int_0^\infty e^{-ru}\lambda(u)du = 1$ implies that $\int_0^\infty e^{-ru}f(u)du = 1/R_0$ which shows how to estimate R_0 from $f(u)$ and r
- The forward generation time density reflects the timing of transmission – we'll use this later

How To Use Early Analysis In Dynamic Model?

$\lambda(a) \equiv$ transmission intensity as function of *age* of infection

$s(t) \equiv$ fraction of the population susceptible to infection at calendar time t

$\pi(t) \equiv$ *incidence* of new infections at calendar time t

$$\pi(t) = s(t)\int_0^{\infty} \pi(t-a)\lambda(a)da \qquad \frac{ds(t)}{dt} = -\pi(t)$$

with initial conditions $s(0)$ and $\pi(a)$ for $a < 0$

- Can modify to account for interventions (Kaplan 2020 *MSOM*), repeat testing/isolation considered downstream

Using The Forward Generation Time

Define $L_F =$ the forward generation time (or lag)

The probability density of L_F equals $f(u) = \lambda(u)/R_0$

The expected lag $E[L_F] = \int_0^\infty u f(u) du \equiv \mu_F$

Rewrite the incidence of infection as

$$\pi(t) = s(t) \int_0^\infty \pi(t-a)\lambda(a)da$$

$$= s(t)R_0 \int_0^\infty \pi(t-a)f(a)da$$

$$= s(t)R_0 E[\pi(t-L_F)]$$

$$\approx s(t)R_0 \pi(t-\mu_F)$$

Final Size in Unmitigated Outbreak

If there is no intervention and an outbreak "runs its course" then the total fraction of the population infected is given by $\phi = \int_0^\infty \pi(u)du$

Consider a susceptible person at time t

Instantaneous hazard of infection, $\gamma(t)$, defined by

$$\gamma(t) = R_0 E[\pi(t - L_F)] \approx R_0 \pi(t - \mu_F)$$

Integrated hazard, Γ, given by

$$\Gamma = \int_0^\infty \gamma(t)dt = R_0 E[\int_0^\infty \pi(t - L_F)dt] \approx R_0 \phi$$

The probability of infection over the outbreak, equivalent to final size, then given by

$$\phi = 1 - e^{-\Gamma} \approx 1 - e^{-R_0 \phi}$$

- If $R_0 = 2.25$, $\phi = 0.85$ (worst case unmitigated outbreak)

Herd Immunity

- If s is the fraction of a population that is susceptible, then *herd immunity* is reached when $R_0 s = 1$ (so each newly infected person infects one person on average)
- From that point in time, each newly infected person no longer "replaces" themself, though infections continue to be transmitted
- Many people erroneously believe that if herd immunity is reached, infections would suddenly stop
- Wrong – *it's all about the journey!*

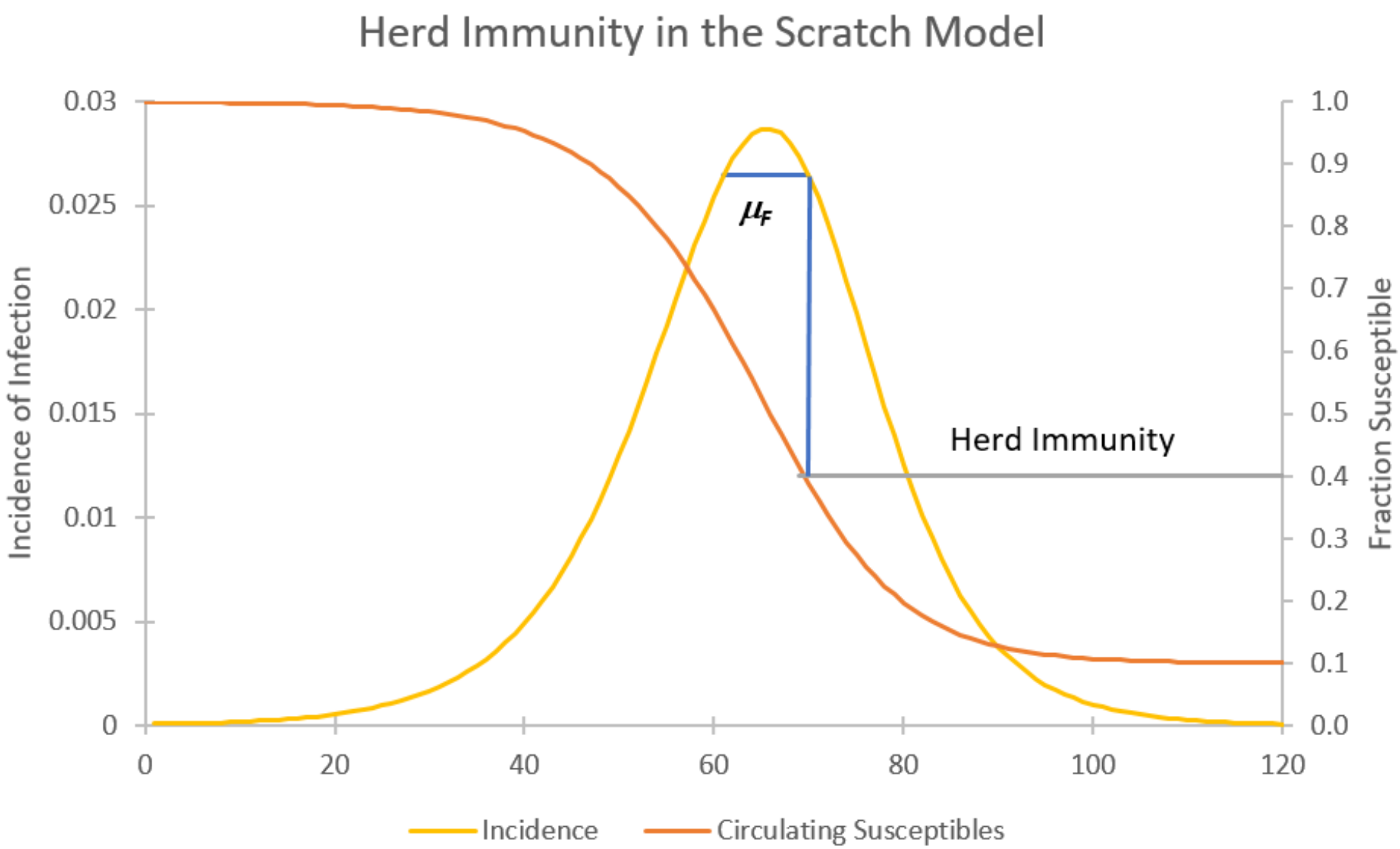
Herd Immunity

Recall $\pi(t) = s(t)R_0E[\pi(t - L_F)] \approx s(t)R_0\pi(t - \mu_F)$

Suppose reach herd immunity at time t^* $(R_0s(t^*) = 1)$

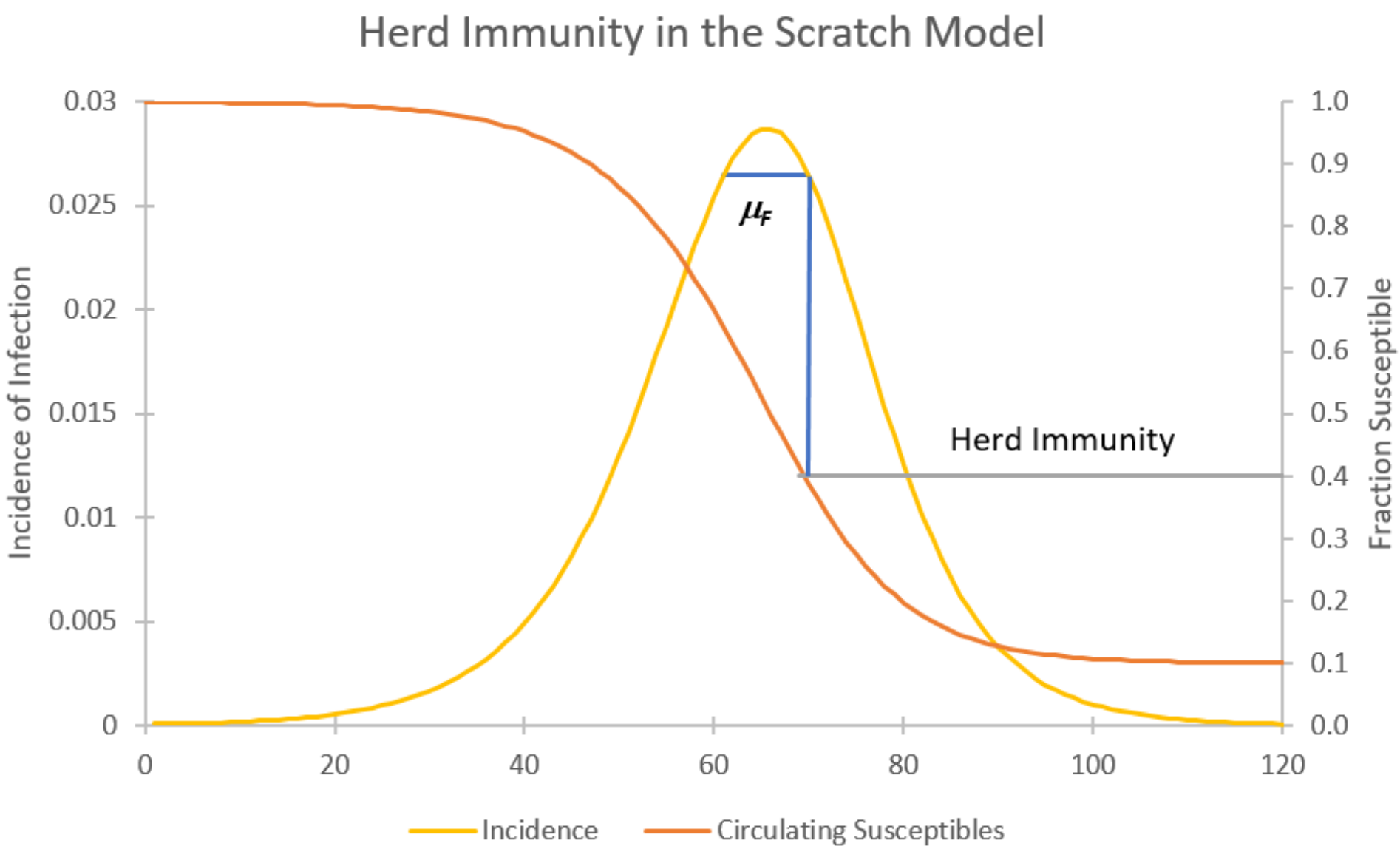
Then $\pi(t^*) \approx s(t^*)R_0\pi(t^* - \mu_F) = \pi(t^* - \mu_F)$

This suggests incidence is maximized at time $t^* - \mu_F/2$ (!!)

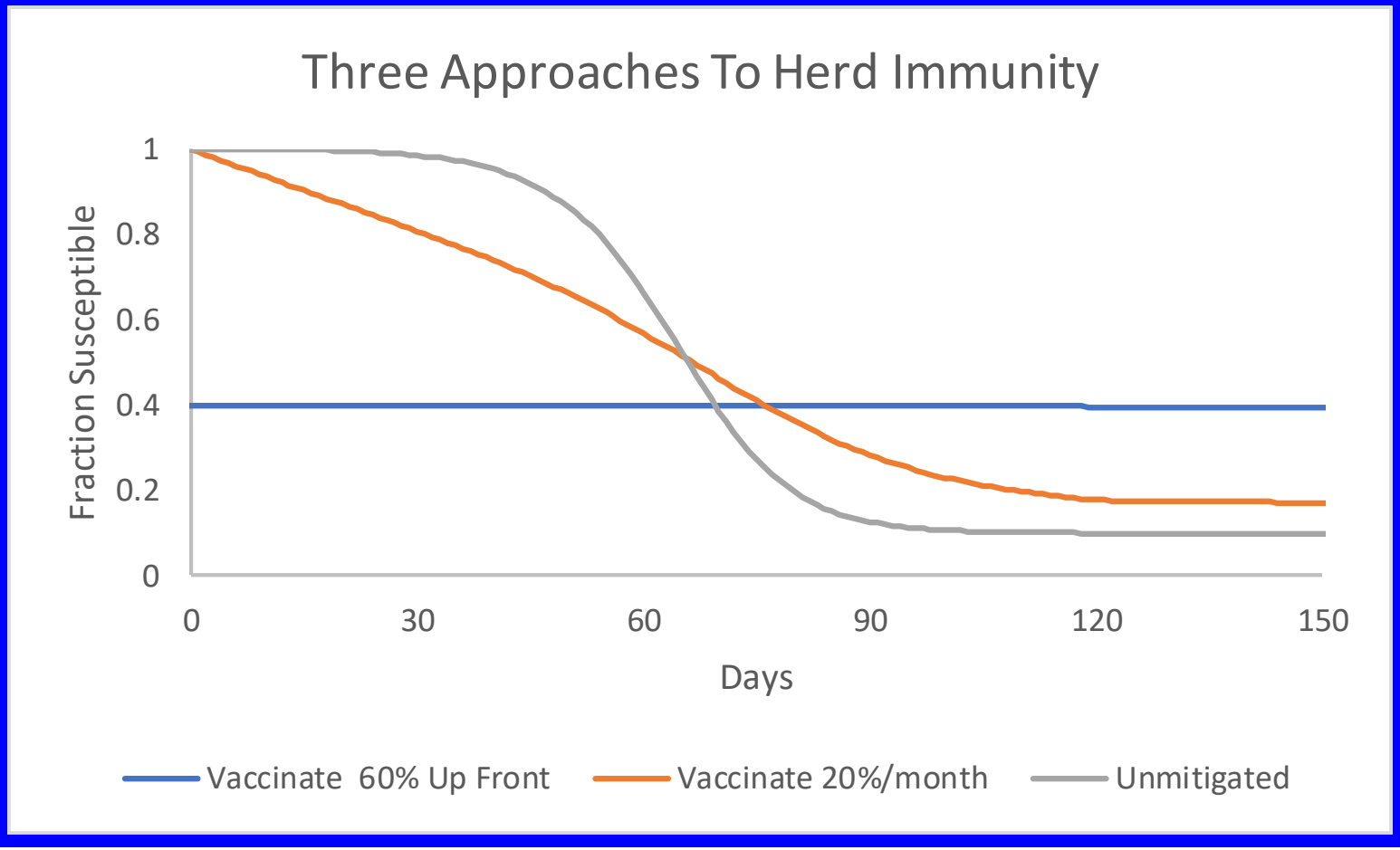

Herd Immunity in the Scratch Model
Incidence of Infection
Fraction Susceptible
0.03
0.025
0.02
0.015
0.01
0.005
0
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
120
μ_F
Herd Immunity
Incidence
Circulating Susceptibles

Herd Immunity Solves an Optimization Problem

- Let $I_L(t) = E[\int_{t-L}^{t} \pi(u)du]$.
- This is the expected number of infections that occurs during the generation interval ending at time *t*.
- Find the maximum of $I_L(t)$. Differentiation with respect to *t* yields the first order condition $\pi(t) = E[\pi(t-L)] \approx \pi(t-\mu_F)$.
- ***This is the condition characterizing the time t^* at which herd immunity is reached!!***
- So, in an unmitigated outbreak, herd immunity is reached at the end of that generation interval containing the maximum expected number of infections.


Herd Immunity in the Scratch Model
Incidence of Infection
Fraction Susceptible
0.03
0.025
0.02
0.015
0.01
0.005
0
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
120
μ_F
Herd Immunity
Incidence
Circulating Susceptibles

Herd Immunity ($R_0 = 2.5$)



Aligning Epidemic Indicators

- We've seen how incidence self-lags via forward generation time

$$\pi(t) = s(t)R_0E[\pi(t-L_F)] \approx s(t)R_0\pi(t-\mu_F)$$

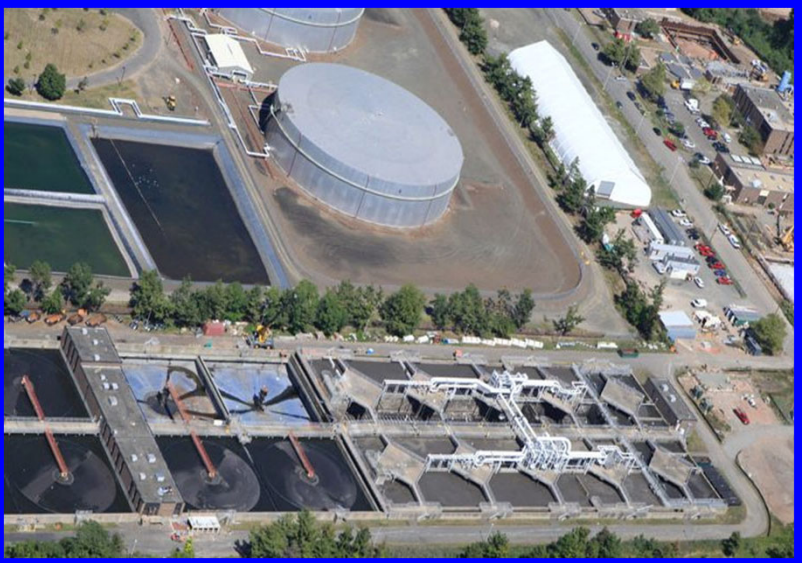
- Suppose $y(t)$ is a (model-scale) lagging indicator of infection like diagnosed cases, hospitalizations, or deaths with characteristic lag L_Y and associated probability density and mean lag; then

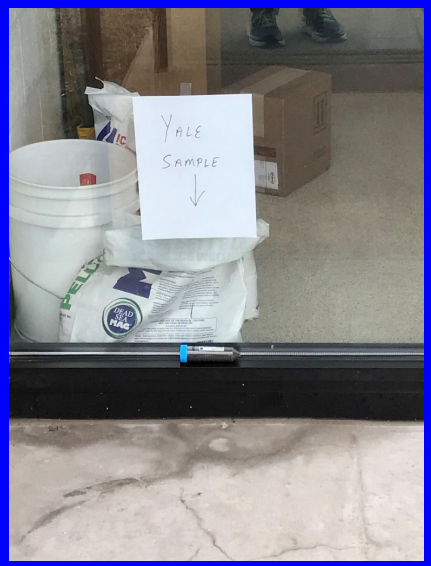
$$y(t) = \int_0^\infty \pi(t-a)f_{L_Y}(a)da = E[\pi(t-L_Y)] \approx \pi(t-\mu_Y)$$

Example: SARS-CoV-2 RNA in Sewage Sludge

- Idea first proposed on March 6 to monitor New Haven outbreak
- Discussion with Prof. Jordan Peccia (Chemical and Environmental Engineering), on March 11; this led to March 16 meeting with other researchers who could contribute (with sample collection, PCR testing, analysis); agreed to focus on local wastewater treatment plant
- First samples were collected on March 19
- Daily sampling continues to date; expanded to other treatment plants in Connecticut (Stamford, Bridgeport, Hartford, etc.)

East Shore Water Pollution Abatement Facility:



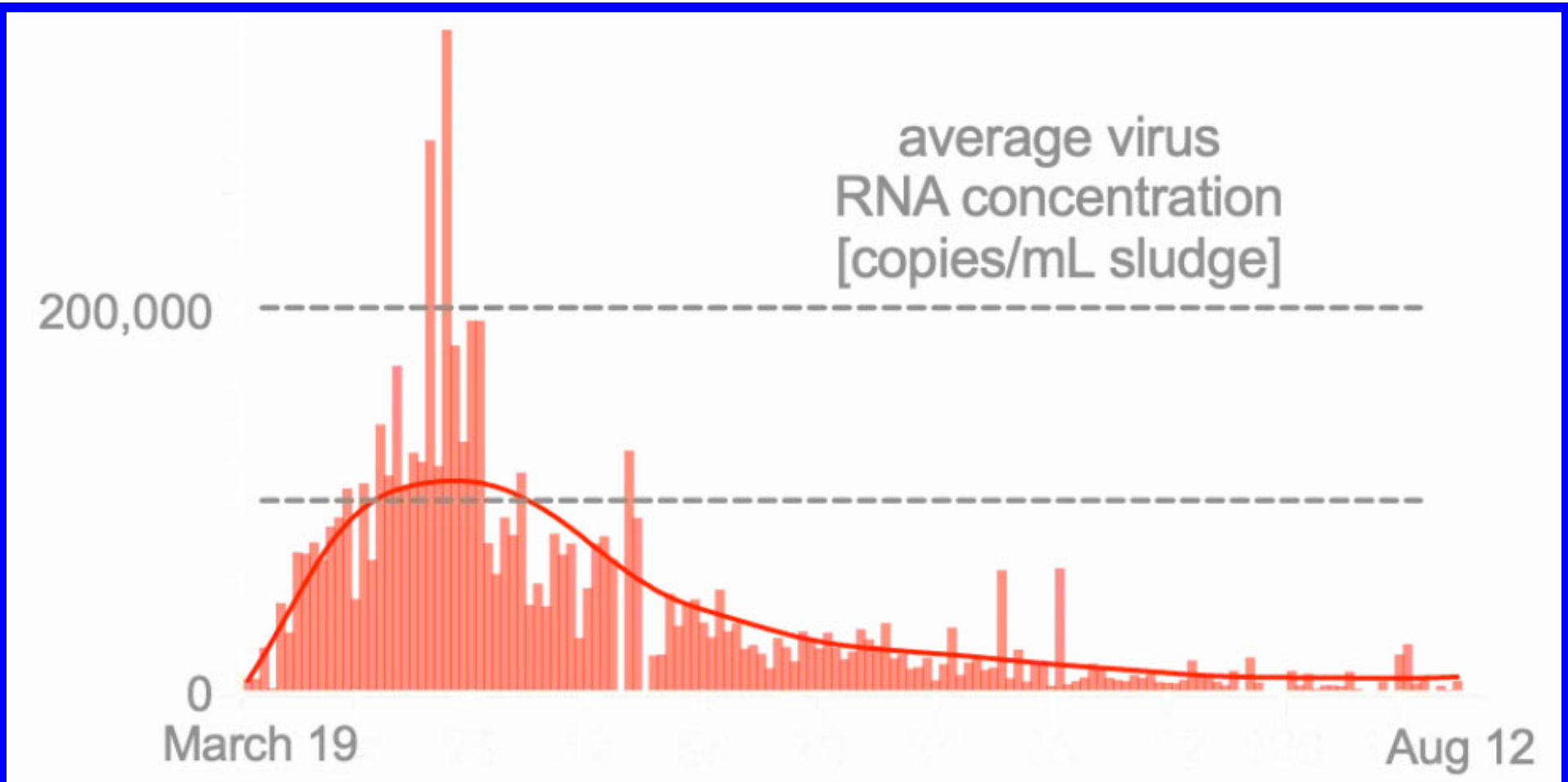


- 40 MGD capacity
- Serves New Haven, CT; East Haven, CT Hamden, CT; Woodbridge, CT
- 4 stage Bardenpho treatment process, 1° and 2° sludge mixed and then enters centrifuges and belt filter press for dewatering along with sludge from treatment plants across CT.
- Dewatered sludge is incinerated.

50 ml samples are collected daily from the mixed 1° and 2° sludge from the East Shore Water Pollution Abatement Facilty only

START: Mar. 19, 2020

Apr. 1, 2020



SARS-CoV-2 RNA concentration in primary sludge at the East Shore Water Pollution Abatement Facility in New Haven, CT

Aligning Sludge Signal with Scratch Model

- Lag from infection until viral detection should follow forward generation time (viral output should grow with $\lambda(t)$, hence $f(t)$)
- This means $L_Y = L_F$ and thus

$$y(t) = E[\pi(t - L_F)] \approx \pi(t - \mu_F)$$

- Model actual sludge viral RNA $Y(t)$ as normal random variable with mean $ky(t)$ and variance $cy(t)$; need to estimate k, c from data

Another Indicator: Hospital Admissions

- Studies of COVID hospitalizations have estimated time from infection to hospitalization by first focusing on incubation time (time from infection to symptoms), and then on time from experiencing symptoms to hospitalization
- Published lag density with mean of 13.5 days (95% probability coverage 4.8 – 27.9 days)
- Model scale hospital admissions represented as

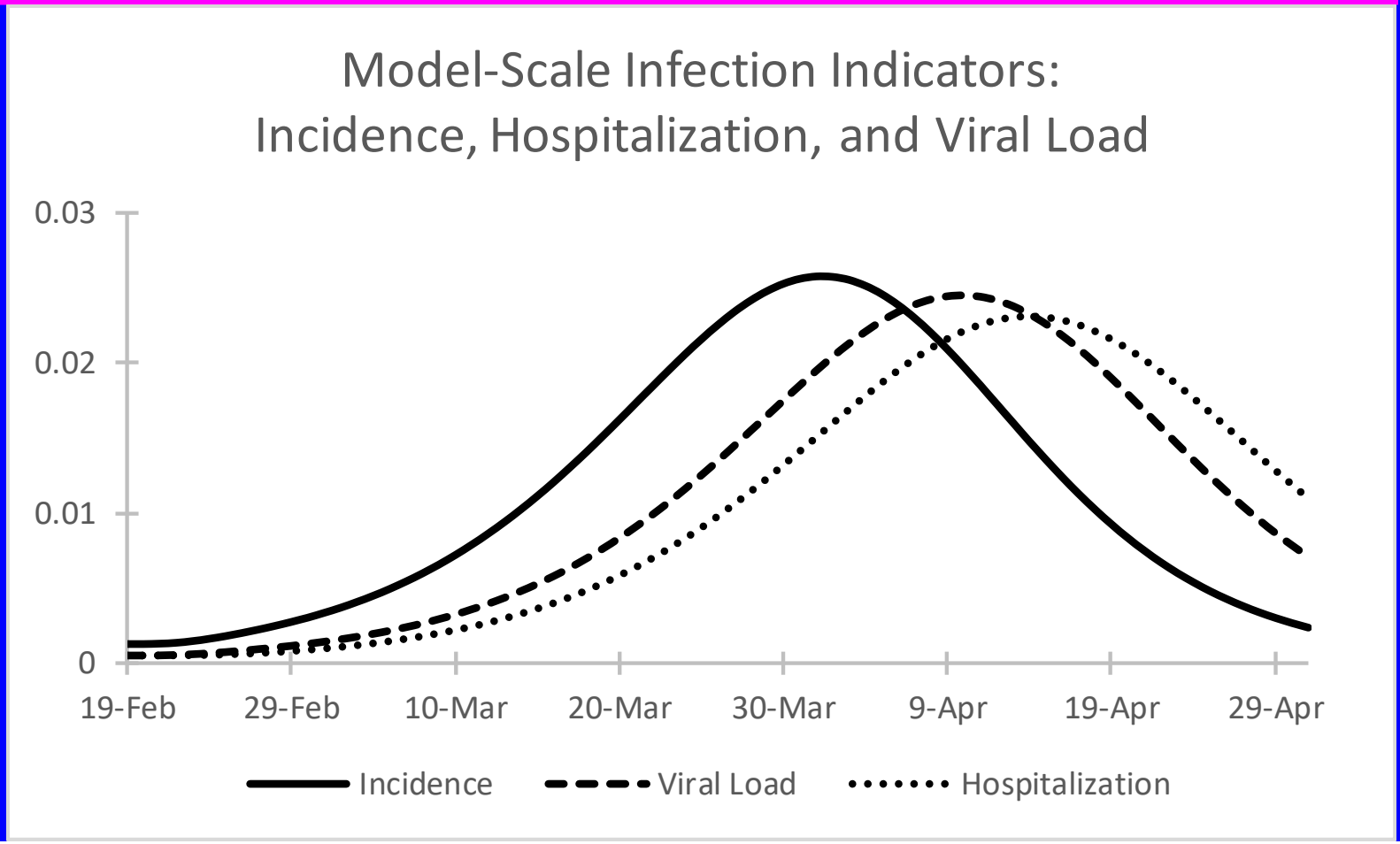
$$h(t) = E[\pi(t - L_H)] \approx \pi(t - \mu_H)$$

- Actual hospital admissions modeled as Poisson random variable with mean and variance of $k_H h(t)$; need to estimate k_H

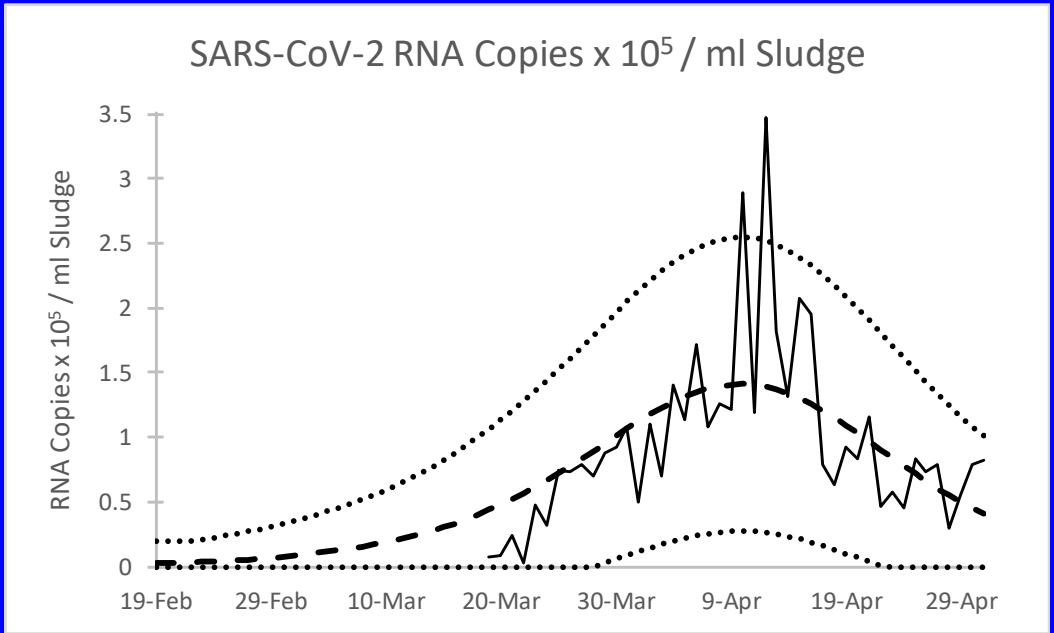
Aligning Epidemic Indicators

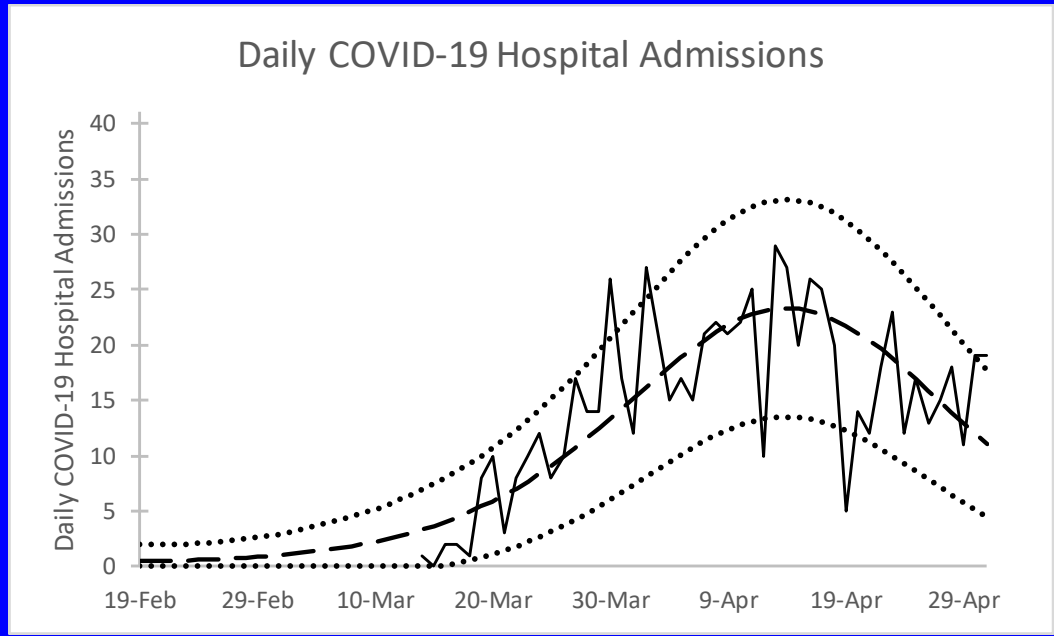
- Given forward generation time and hospitalization lag densities from epidemiological studies, estimate three scaling constants, R_0 and cumulative incidence up as of February 19 (initial condition)
- R_0 measures strength of the outbreak; initial condition places the epidemic wave (higher initial condition pulls outbreak earlier in time; smaller initial condition pushes it later)

Results on the Model Scale



Empirical Results





- Estimated $R_0 = 2.38$ (std error 0.10); cumulative incidence as of February 19 = 0.016 (std error 0.003)
- Suggests sludge signal leads hospital admissions by $\mu_H - \mu_F = 4.6$ days

Different Problem: Repeat Screening to Detect and Control SARS-CoV-2 Transmission

- Detect and isolate new infections through viral testing
- Gain control of transmission; prevent ignition of local outbreaks
- Detecting new infections requires intensive screening
- Focus on screening for asymptomatic infections to shorten time from infection to isolation

JAMA Network

JAMA Health Forum

Logistics of Aggressive Community Screening for Coronavirus 2019

Edward H. Kaplan, PhD[1,2,3]; Howard P. Forman, MD, MBA[1,2,4]

Preventing Transmission via Isolation

- Suppose an infected person is *isolated* at T time units after infection

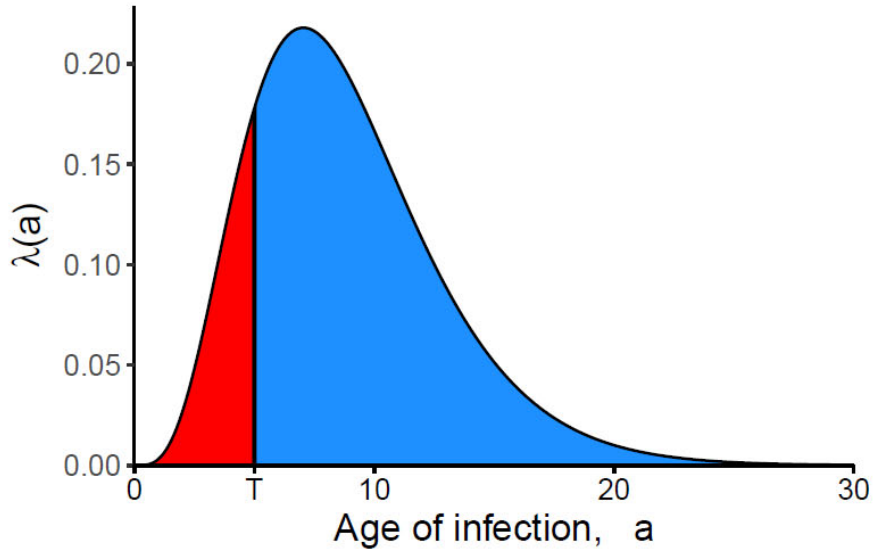
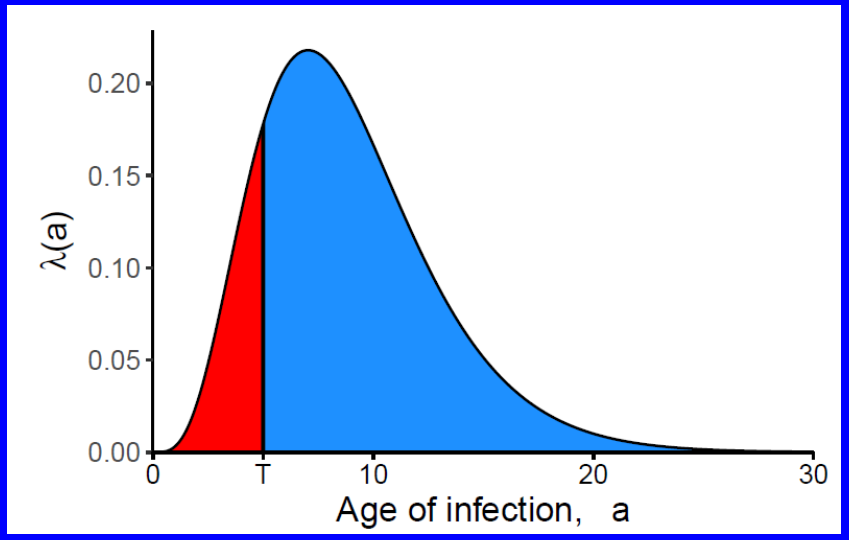


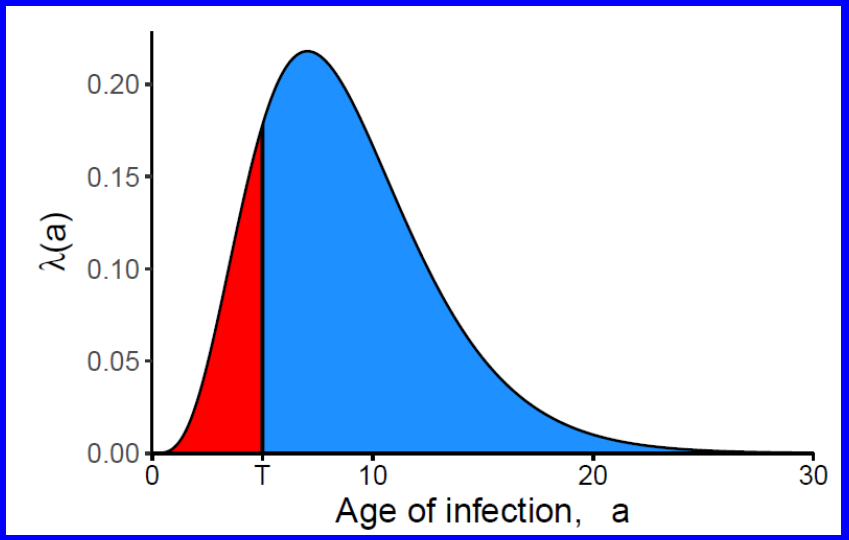
Figure 1: Impact of isolation. For a person isolated at a random time T after infection, the blue shaded area shows the expected number of further infections whose transmissions are prevented by the isolation, and the red area shows the expected number of further infections that escape isolation and are still transmitted.

Preventing Transmission via Isolation



- *T* is a random variable
- The distribution of *T* depends upon how infected persons are detected
  - Test based on self-identification of symptoms relates *T* to *incubation time* and test sensitivity
  - Asymptomatic screening relates *T* to testing frequency and evolution of test sensitivity with time from infection
  - Test based on contact tracing relates *T* to forward/backward generation time and evolution of test sensitivity

Preventing Transmission via Isolation



- Expected transmission at age a of infection, $\lambda(a)$, only experienced if $T > a$
- Effectively reduces transmission rate from $\lambda(a)$ to $\lambda(a)\Pr\{T > a\}$
- Reduces reproductive number R_0 to R_T where

$$R_T = \int_0^\infty \lambda(a)\Pr\{T > a\}da < \int_0^\infty \lambda(a)da = R_0$$

- Possible for $R_T < 1$, showing how isolation could end outbreak

Repeat Testing in Fixed Population

(e.g. Residential College)

- Perfect test, random screening with mean intertest time τ days

$$\Pr\{T > a\} = e^{-a/\tau} \text{ for } a > 0$$

- Perfect test, scheduled screening once every τ days

$$\Pr\{T > a\} = \max(1 - a/\tau, 0) \text{ for } a > 0$$

- Scheduled more efficient than random since $e^{-a/\tau} > 1 - a/\tau$
- Imperfect repeat test every τ days with test sensitivity σ; T is now time to first positive test

$$\Pr\{T > a\} = (1-\sigma)^{\lfloor \frac{a}{\tau} \rfloor} \left(1 - \sigma \frac{a - \lfloor \frac{a}{\tau} \rfloor \tau}{\tau}\right) \text{ for } a > 0.$$

Age-of-Infection Dependent Sensitivity

- Let $\sigma(a)$ = test sensitivity at a time units after infection
- Assume tests at different times are independent; scheduled repeat testing with intertest interval τ
- Let $g_T(a)$ denote probability density of T, time to first positive test

$$g_T(a) = \frac{\sigma(a)}{\tau} \prod_{k=1}^{\lfloor \frac{a}{\tau} \rfloor} (1 - \sigma(a - k\tau)) \text{ for } a > 0$$

$$\Pr\{T > a\} = \int_a^\infty g_T(u)du$$

Age-of-Infection Dependent Sensitivity

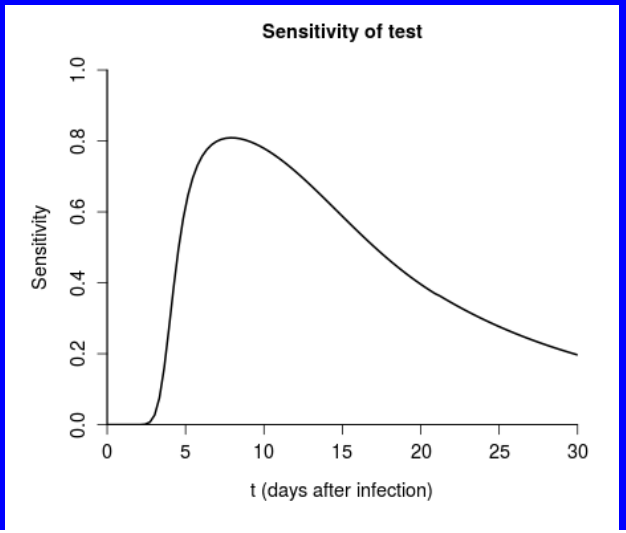
- Box function: sensitivity = 0 for $a < w$ and $a > r$; otherwise = σ
  - w is silent test *window*; r is test *reach*

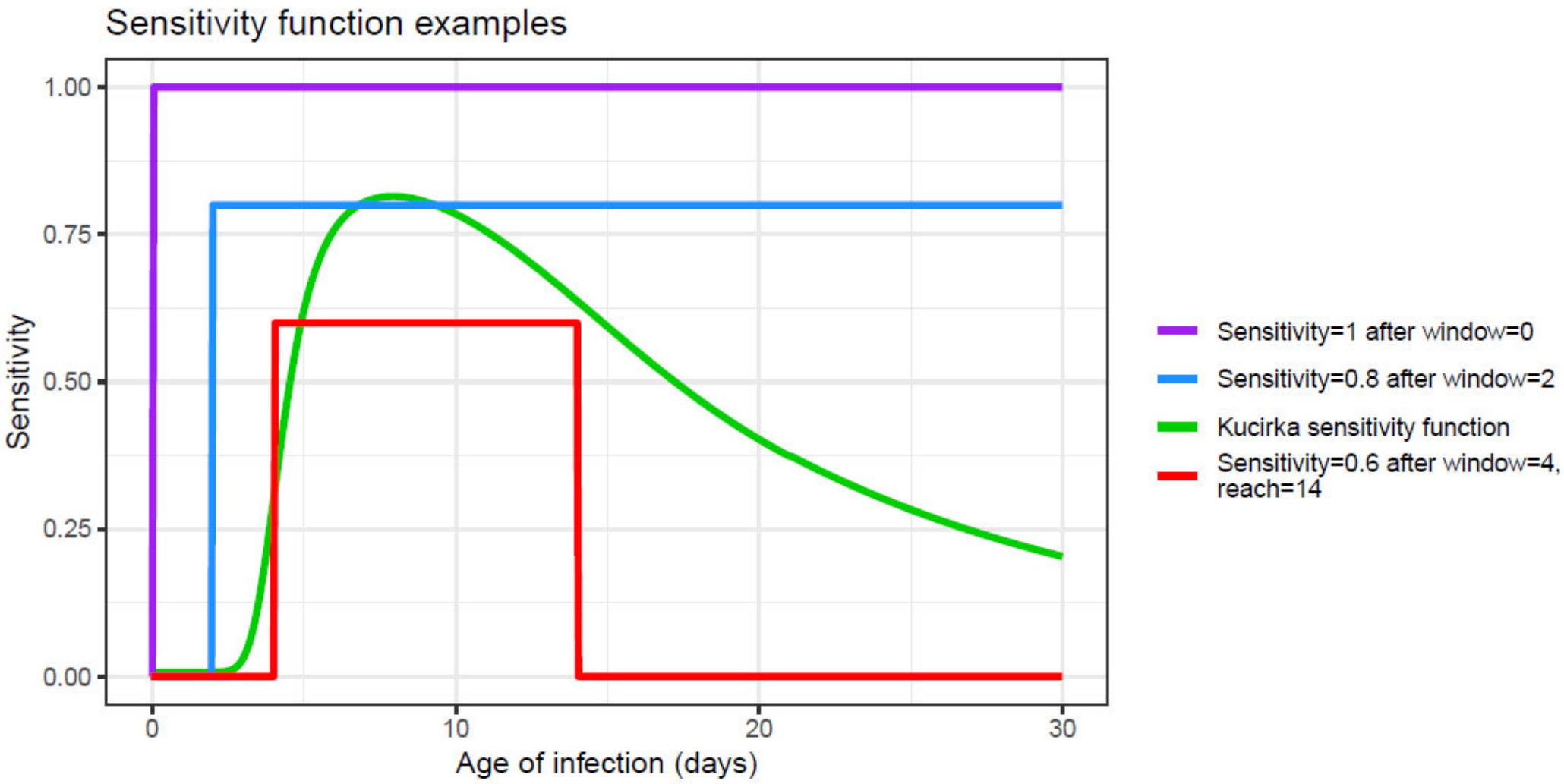
$$\sigma(a) = \sigma 1_{\{w<a<r\}}$$

- Empirical distribution based on empirical study testing same infected people daily over time from symptoms; impute time before symptoms (Kucirka *et al* 2020)

$$\sigma(a) = \begin{cases} \text{logistic}\left(-29.966 + 37.713\log(a) - 14.452(\log a)^2 + 1.721(\log a)^3\right) & 0 \le a \le 21 \\ \text{logistic}\left(6.878 - 2.436\log(a)\right) & a > 21 \end{cases}$$

$$\text{logistic}(z) = e^z/(1+e^z)$$




Sensitivity function examples
1.00
0.75
0.50
0.25
0.00
Sensitivity
0
10
20
30
Age of infection (days)
Sensitivity=1 after window=0
Sensitivity=0.8 after window=2
Kucirka sensitivity function
Sensitivity=0.6 after window=4, reach=14

Isolation Delay

- It takes time from when a test is taken until the test is processed and the infected person is notified and isolated; let ℓ be isolation delay
- Suppose T is the time from infection until the date of the first positive test; T_ℓ is the time from infection until isolation

$$\Pr\{T_\ell > a\} = \begin{cases} 1 & 0 \le a < \ell \\ \Pr\{T_0 > a - \ell\} & a \ge \ell. \end{cases}$$

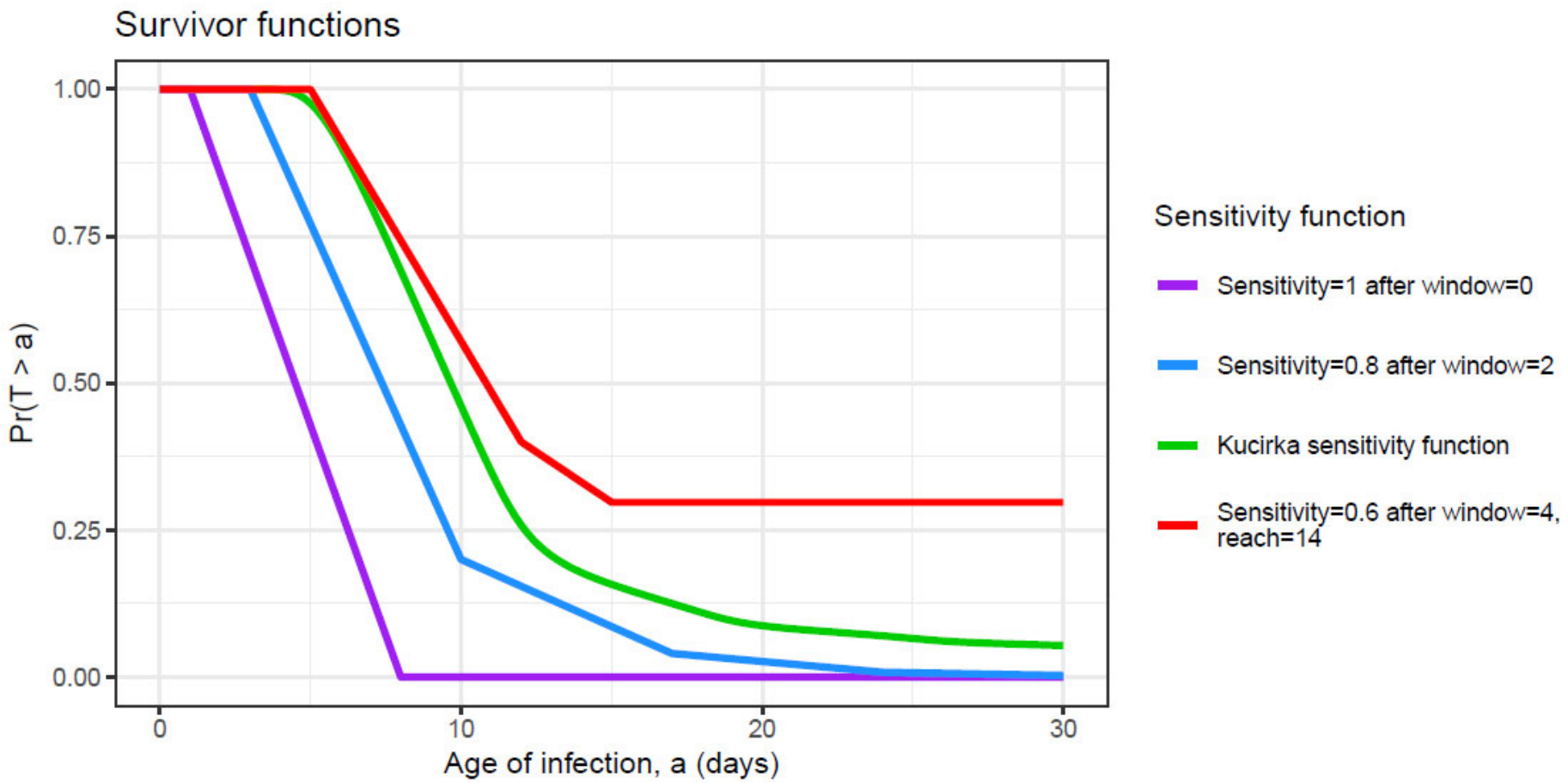


Figure 3: Probability that the time from infection to isolation exceeds a. Here the isolation delay in all four scenarios was taken to be 1 day.

Combining Transmission and Time to Isolation

- Recall transmission intensity given by $\lambda(a)\Pr\{T > a\}$

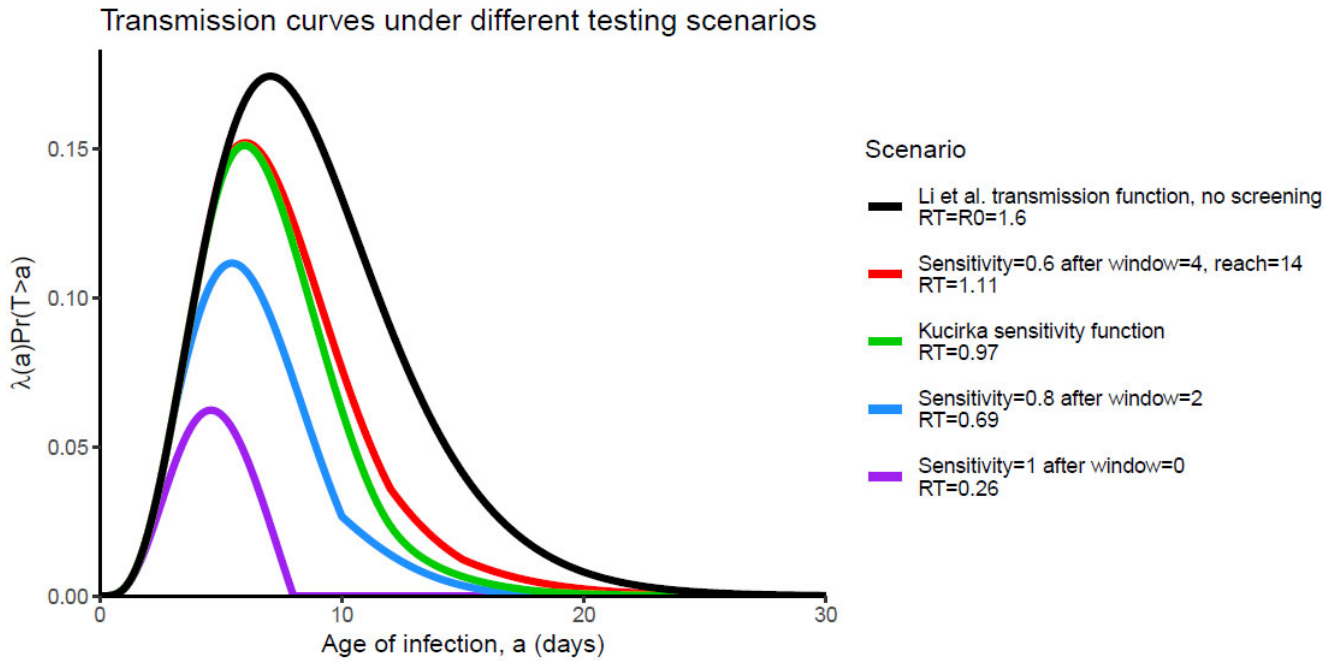


Figure 4: Transmission curves under different testing scenarios.

Putting It All Together

- Suppose that in addition to internally generated infections, population members are exposed to *imported* infections at rate $v(t)$, meaning in addition to internal transmission, $v(t)s(t)$ infections are imported at time t (think of v for *v*isitor)
- Repeat testing scratch model: from initial conditions (gateway testing)

$$\pi(t) = s(t)\left\{\int_0^\infty \pi(t-a)\lambda(a)\Pr\{T > a\}\,da + v(t)\right\} \quad \text{for } t > 0$$

$$\frac{ds(t)}{dt} = -\pi(t)$$

Repeat Screening for SARS-CoV-2

- Easy to determine true and false positive detection rates, fraction of the population isolated at any time, etc. from model
- Isolation Rate for True Positives: $\delta_{TP}(t) = \int_{a=0}^{\infty} \pi(t-a) g_T^{(t-a)}(a)\, da$
- Isolation Rate for False Positives: $\delta_{FP}(t) = s(t-\ell)\phi/\tau$
- Number in isolation at time t (duration of isolation = Δ

$$\iota(t) = \int_{\max(0,t-\Delta)}^{t} \delta(u) du$$

Illustration

- Model available at https://jtwchang.shinyapps.io/testing/
- All the features described above are there
- Criterion: find most difficult scenarios that keep cumulative infections < 5% over course of 80 day semester (i.e. this fall!)
- For different choices of forward generating time density $f(t)$, test sensitivity function, imported infection rate, initial conditions, isolation delay, find largest R_0 that keeps infections < 5% over term as a function of scheduled testing frequency

Generating Times and Sample Trajectories

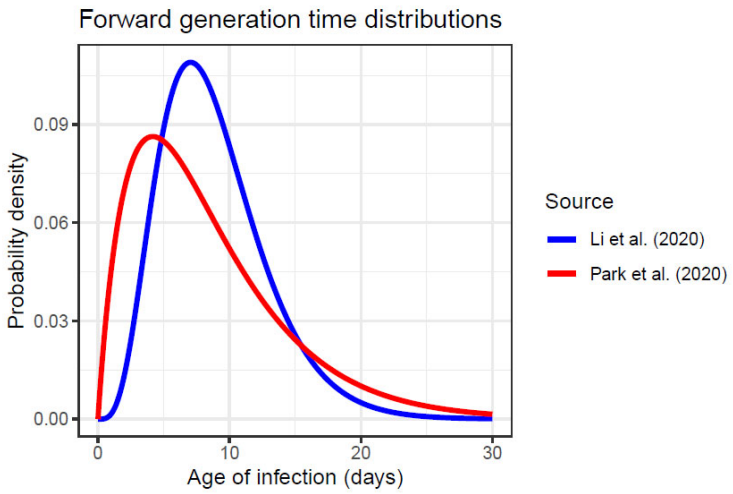


Figure 6: Two estimated generation time distributions found in published studies. We refer to these distributions as featuring relatively *early transmission* (Park et al. 2020) and *late transmission* (Li et al. 2020).

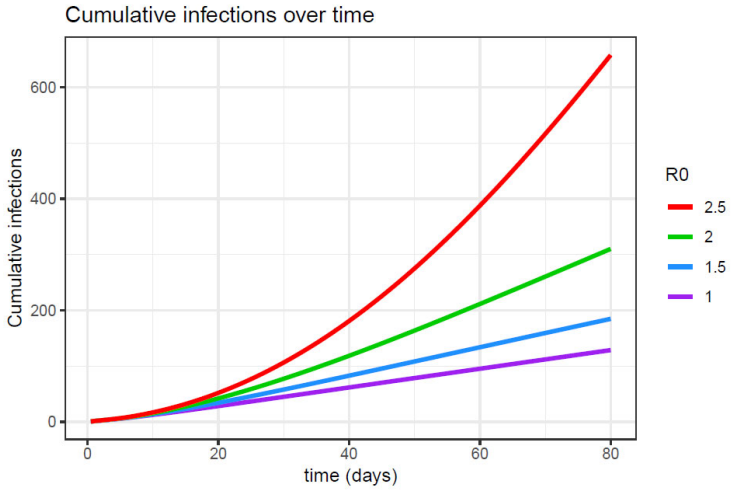


Figure 7: Cumulative infections over time in scenarios with testing every 3 days and fixed infectivity function, sensitivity function, and delays, as R_0 varies.

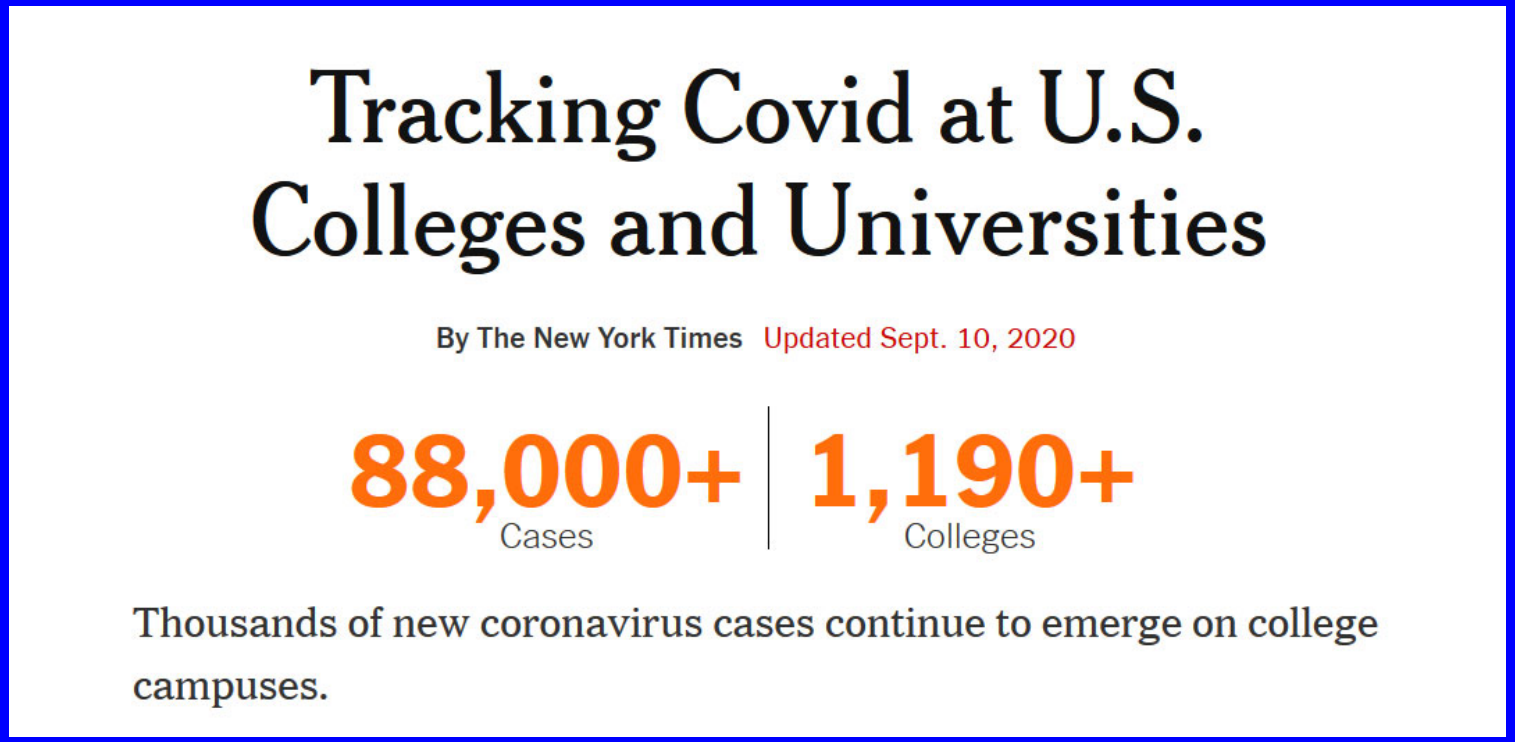
Results: Weekly Testing

Scenario $(f(a), \sigma(a))$	Maximal R_0	Infections	Average Isolated	Maximum Isolated	Positive Tests/day
Li et al. Kucirka	1.6	472	87	152	7
Li et al. step-function	2.25	465	93	155	8
Park et al. Kucirka	1.4	447	87	139	7
Park et al. step-function	1.8	456	99	156	8

Results: Testing Every 3 Days

Scenario $(f(a), \sigma(a))$	Maximal R_0	Infections	Average Isolated	Maximum Isolated	Positive Tests/day
Li et al. Kucirka	2.3	474	143	207	11
Li et al. step-function	4.8	491	150	206	12
Park et al. Kucirka	1.75	459	143	194	11
Park et al. step-function	2.65	499	153	197	12

COVID College Outbreaks



Tracking Covid at U.S. Colleges and Universities

By The New York Times Updated Sept. 10, 2020

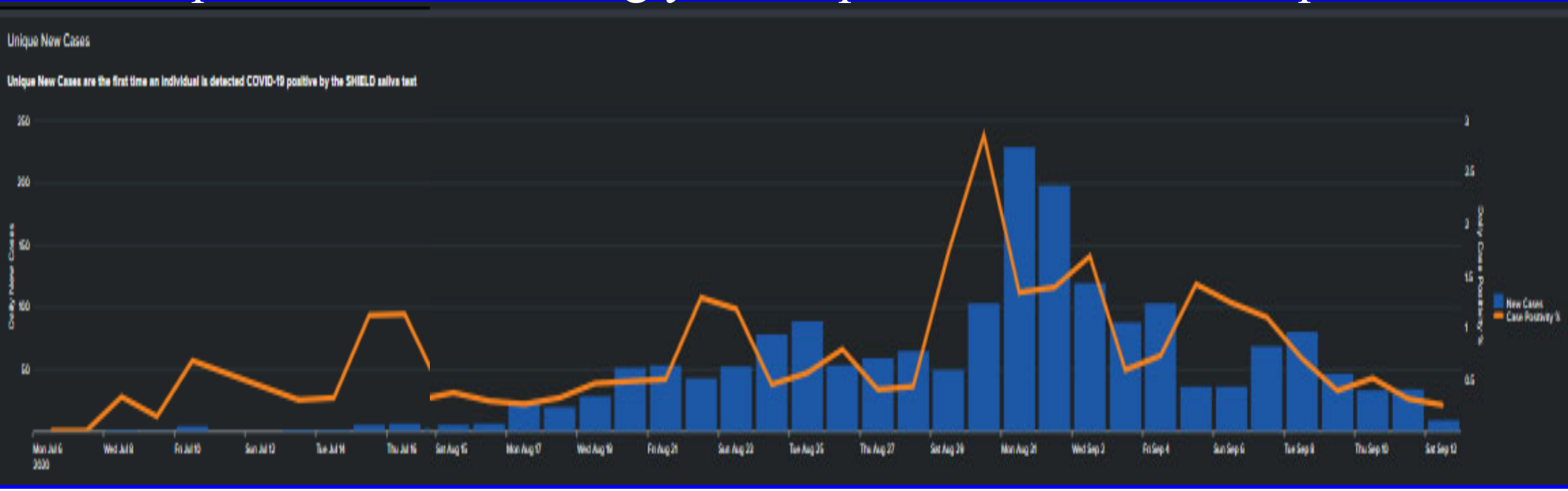
88,000+ Cases | **1,190+** Colleges

Thousands of new coronavirus cases continue to emerge on college campuses.

- Most of the colleges in question have not implemented repeat testing

Cautionary Note: University of Illinois

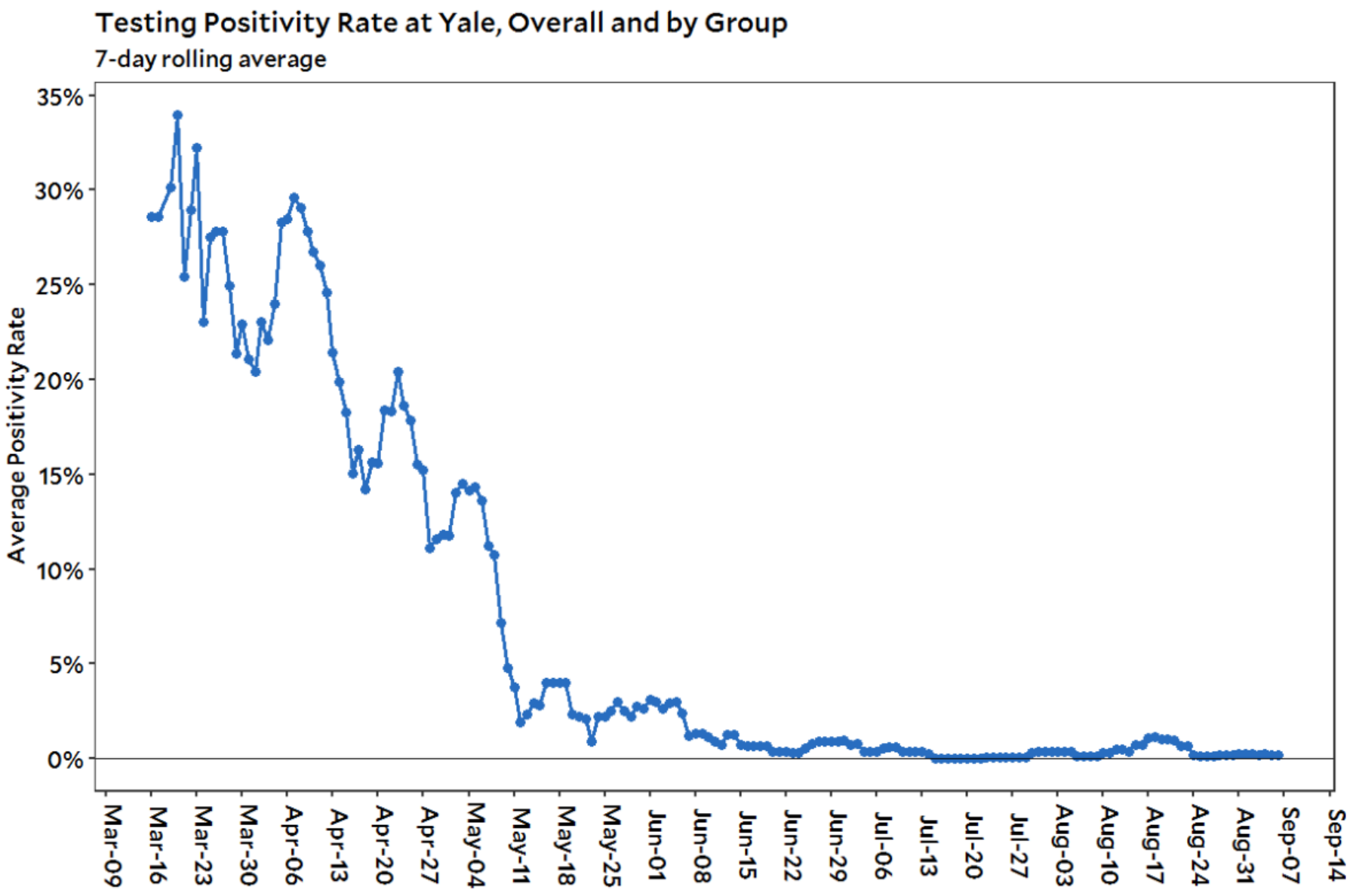
- Twice weekly testing on campus
- Large outbreak last week of August
- Some persons who knowingly tested positive still attended parties etc



First Two Weeks Look Good

Daily Testing Data

Population Type	Sep 09	Sep 10	Sep 11	Sep 12	Sep 13	Sep 14	Sep 15	Total Tests in Last 7 Days	Individuals Tested in Last 7 Days
Students									
# Tests	**304**	**1364**	**1446**	**148**	**4**	**1404**	**1442**	**8144**	**4931**
# Positives	**1**	**0**	**2**	**0**	**0**	**0**	**0**		**3**
Undergraduates living on campus									
# Tests	47	611	644	21	1	648	649	3490	1785
# Positives	0	0	0	0	0	0	0		0
Undergraduates living off campus									
# Tests	59	496	538	35	2	513	530	2825	1531
# Positives	0	0	2	0	0	0	0		2
Graduate & Professional									
# Tests	198	257	264	92	1	243	263	1829	1615
# Positives	1	0	0	0	0	0	0		1
Faculty & Staff									
# Tests	**327**	**130**	**111**	**34**	**0**	**108**	**141**	**1177**	**1119**
# Positives	**0**	**0**	**0**	**0**	**0**	**0**	**0**		**0**
# Tests	**631**	**1494**	**1557**	**182**	**4**	**1512**	**1583**	**9321**	**6050**
# Positives	**1**	**0**	**2**	**0**	**0**	**0**	**0**		**3**


Testing Positivity Rate at Yale, Overall and by Group
7-day rolling average
Average Positivity Rate
35%
30%
25%
20%
15%
10%
5%
0%
Mar-09
Mar-16
Mar-23
Mar-30
Apr-06
Apr-13
Apr-20
Apr-27
May-04
May-11
May-18
May-25
Jun-01
Jun-08
Jun-15
Jun-22
Jun-29
Jul-06
Jul-13
Jul-20
Jul-27
Aug-03
Aug-10
Aug-17
Aug-24
Aug-31
Sep-07
Sep-14

References

- Containing 2019-nCoV (Wuhan) coronavirus, *HCMS* 23, 311–314 (https://doi.org/10.1007/s10729-020-09504-6S)
- COVID-19 Scratch Models to Support Local Decisions, *MSOM* 22, 645 – 655 (https://doi.org/10.1287/msom.2020.0891)
- SARS-CoV-2 RNA concentrations in primary municipal sewage sludge as a leading indicator of COVID-19 outbreak dynamics (J Peccia *et al*), in press *Nature Biotechnology* (https://www.medrxiv.org/content/10.1101/2020.05.19.20105999v2)
- Aligning SARS-CoV-2 Indicators via an Epidemic Model: Application to Hospital Admissions and RNA Detection in Sewage Sludge (with J Peccia *et al,* https://www.medrxiv.org/content/10.1101/2020.06.27.20141739v3)
- Logistics of Aggressive Community Screening for Coronavirus 2019 (with H Forman), *JAMA Health Forum* https://jamanetwork.com/channels/health-forum/fullarticle/2765693
- Repeat SARS-CoV-2 Testing Models for Residential College Populations (JT Chang, FW Crawford, EH Kaplan, https://www.medrxiv.org/content/10.1101/2020.07.09.20149351v2)