medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Aligning SARS-CoV-2 Indicators via an Epidemic Model: Application to Hospital Admissions and RNA Detection in Sewage Sludge

Edward H. Kaplan*, Dennis Wang†, Mike Wang‡,
Amyn A. Malik§, Alessandro Zulli¶, Jordan Peccia‖

August 2020

Abstract

Ascertaining the state of coronavirus outbreaks is crucial for public health decision-making. Absent repeated representative viral test samples in the population, public health officials and researchers alike have relied on lagging indicators of infection to make inferences about the direction of the outbreak and attendant policy decisions. Recently researchers have shown that SARS-CoV-2 RNA can be detected in municipal sewage sludge with measured RNA concentrations rising and falling suggestively in the shape of an epidemic curve while providing an earlier signal of infection than hospital admissions data. The

*William N. and Marie A. Beach Professor of Operations Research, Professor of Public Health, Professor of Engineering, Yale School of Management, 165 Whitney Avenue, New Haven 06511, CT USA. email: edward.kaplan@yale.edu

†Yale School of Medicine, New Haven, CT, USA.

‡Yale School of Medicine, New Haven, CT, USA.

§Yale School of Medicine, Yale Institute for Global Health, New Haven, CT, USA.

¶Department of Chemical and Environmental Engineering, School of Engineering and Applied Science, Yale University, New Haven 06511, CT, USA.

‖Thomas E. Golden, Jr. Professor of Chemical & Environmental Engineering, Department of Chemical and Environmental Engineering, School of Engineering and Applied Science, Yale University, New Haven 06511, CT USA. email: jordan.peccia@yale.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

present paper presents a SARS-CoV-2 epidemic model to serve as a basis for estimating the incidence of infection, and shows mathematically how modeled transmission dynamics translate into infection indicators by incorporating probability distributions for indicator-specific time lags from infection. Hospital admissions and SARS-CoV-2 RNA in municipal sewage sludge are simultaneously modeled via maximum likelihood scaling to the underlying transmission model. The results demonstrate that both data series plausibly follow from the transmission model specified, provide a direct estimate of the reproductive number $R_0 = 2.38$, and suggest that the detection of viral RNA in sewage sludge leads hospital admissions by 4.6 days on average.

Keywords: SARS-CoV-2, COVID-19, epidemic indicators, wastewater epidemiology, sewage sludge viral RNA concentration, COVID-19 hospital admissions, probability model

1 Introduction

Ascertaining the state of coronavirus outbreaks is crucial for public health decision-making. Absent repeated representative viral test samples in the population (Kaplan and Forman 2020), public health officials and researchers alike have relied on lagging indicators of infection to make inferences about the direction of the outbreak and attendant policy decisions. How useful these indicators are depends upon their typical lags behind the incidence of infection. Some indicator lags, such as time from infection to hospitalization, have been studied empirically (Lewnard et al 2020, CDC 2020, MIDAS 2020). Other indicators have been proposed with the hope that they would greatly reduce the lag time from infection. One such promising indicator is measured SARS-CoV-2 RNA concentration in municipal wastewater (Hart and Halden

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2020, Peccia et al 2020). How much earlier might such a signal inform officials of changes in the state of the outbreak?

This paper tackles this question by using an epidemic transmission model to create model-scale versions of whatever indicator is of interest, and then scales these model quantities to match observed indicator values in the real world. This approach clarifies the time lags that should be expected from SARS-CoV-2 incidence to whichever indicator is of interest, and by doing so makes it possible to compare the relative timing of one indicator to another, providing the model fit to the data is sufficiently close.

Our study takes advantage of recently conducted research tracking the local SARS-CoV-2 outbreak in the New Haven, Connecticut, USA metropolitan area. As reported by Peccia et al (2020), daily SARS-CoV-2 RNA concentrations were obtained by sampling sewage sludge from the local wastewater treatment plant and conducting PCR tests to determine virus RNA concentration. Daily COVID-19 admissions to the Yale New Haven Hospital restricted to residents of the same four towns served by this wastewater treatment plant were also recorded over the same time period. An epidemic model developed by Kaplan (2020b) was taken as the basis for calibrating these two lagging indicators while simultaneously estimating the initial condition and reproductive number R_0 of this outbreak. This paper details the methodology employed and results obtained from doing so.

The next section presents a quick description of the transmission model reported in Kaplan (2020b). In Section 3, a simple method is described for

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

linking model-scale lagging epidemic indicators to SARS-CoV-2 incidence based on the model and appropriately defined lag probability density functions, which enables a model-scale comparison of different indicators to see how they should appear over the course of an outbreak (Figure 1). Section 4 presents a simple statistical approach to analyzing real-world indicator data by scaling modeled indicators up to observed values based on maximum likelihood estimation while also estimating the initial condition and reproductive number of the epidemic wave from the underlying transmission model. We simultaneously scale hospital admissions and the RNA virus concentration observed in the sewage sludge to the epidemic model (Table 1, Figure 2 and 3). The results show that accounting for the inherent noise in the data, both the virus RNA concentration in the sewage sludge and hospital admissions match the model expectations reasonably well. The analysis verifies that there is about a 4.6 day separation between the sludge RNA concentration and hospital admissions curve, consistent with earlier analysis based on statistical time-series analysis (Peccia et al 2020). Section 5 provides a summary of the key points of the paper.

2 Transmission Model

Data detailing person-to-person SARS-CoV-2 transmission in Wuhan were reported by Li et al (2020). These data enabled an early model-based assessment of prospects for containing coronavirus via isolation and quarantine

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(Kaplan 2020a), while that analysis was extended to a dynamic transmission model for SARS-CoV-2 transmission in Connecticut (Kaplan 2020b). This latter model incorporates infection-age-dependent transmission, and thus falls into the class of renewal equation epidemic models (Heesterbeek and Dietz 1996, Champredon and Dushoff 2015). The key model element is the age-of-infection dependent transmission rate $\lambda(a)$, which can be thought of as the instantaneous transmission intensity of an individual who has been infected for a time units. At the beginning of an outbreak when an infectious person is embedded in an otherwise susceptible population, the expected number of infections transmitted per infectious person equals the reproductive number R_0, which is given by

$$R_0 = \int_0^\infty \lambda(a)da \tag{1}$$

as is well known. Li et al (2020) reported estimates of both the exponential growth rate r and backwards generation time probability density function $b(a)$, enabling $\lambda(a)$ to be written as

$$\lambda(a) = e^{ra}b(a),\ a > 0 \tag{2}$$

(Kaplan 2020a, 2020b; Britton and Tomba 2019; Champredon and Dushoff 2015; Wallinga and Lipsitch 2007), which together imply a point estimate of $R_0 = 2.26$ (Kaplan 2020a), consistent with values widely reported elsewhere (Ferguson et al 2020, Kissler et al 2020, MIDAS 2020, Park et al 2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

An alternative representation of $\lambda(a)$ is

$$\lambda(a) = R_0 f(a) \tag{3}$$

where $f(a)$ is the forward generation time density that dictates the timing of transmission (Britton and Tomba 2019; Champredon and Dushoff 2015; Wallinga and Lipsitch 2007). We adopt this representation in the present analysis, as it enables estimation of the underlying reproductive number R_0 directly from the data at our disposal.

The transmission model developed in Kaplan (2020b) that will be used to anchor our infection indicators analysis follows. Let

$\psi(t) \equiv$ transmission potential (or force of infection) at chronological time t;

$s(t) \equiv$ fraction of the population that is susceptible to infection at chronological time t;

$\pi(a,t) \equiv$ fraction of the population that has been infected for duration a at time time t;

$\pi(0,t) =$ incidence of infection at time t.

Given initial conditions $\pi(a,0)$ and $s(0)$ the model equations are:

$$\psi(t) = \int_0^\infty \lambda(a)\pi(a,t)da, \;\; t > 0 \tag{4}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$\pi(0,t) = s(t)\psi(t),\ t > 0 \tag{5}$$

$$\frac{ds(t)}{dt} = -\pi(0,t),\ t > 0 \tag{6}$$

$$\pi(a,t) = \begin{cases} \pi(a-t,0) & 0 < t \leq a \\ \\ \pi(0,t-a) & 0 < a \leq t \end{cases} \tag{7}$$

Equation (4) defines the transmission potential at time t, which is infection-age-dependent transmission $\lambda(a)$ weighted by the infection-age-dependent prevalence of infection in the population $\pi(a,t)$; equation (5) equates SARS-CoV-2 incidence to the product of the fraction of the population that is susceptible and the transmission potential; equation (6) depletes susceptibles with the incidence of infection; and equation (7) aligns the fraction of the population infected for duration a at time t with the incidence of infection at time $t-a$, adjusting for the initial conditions at time zero. There is no additional accounting for the duration of infectiousness because the time course of infection is already built into $\lambda(a)$.

The transmission function employed for the present analysis is given by equation (3) using the forward generation time density $f(a)$ implied by Li et al (2020), which is a gamma density with mean (standard deviation) equal to

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

8.86 (4.02) days[1]. The reproductive number R_0 and initial condition $\pi(a, 0)$ are estimated from the data as described below.

3 Transmission Indicators and Time Lags

In the absence of repeated representative viral testing in a population, officials and researchers alike have turned to lagging indicators of infection such as diagnosed COVID-19 cases, hospitalizations, and deaths to monitor the state of the outbreak. How useful such indicators are depends upon their lag time from infection. Let $y(t)$ be the value of a model-scale infection indicator that represents a distributionally lagged signal of the incidence of infection. Specifically, denote L_Y as the time lag from infection, and define $f_{L_Y}(a)$ as the probability density function governing the lag L_Y. The model-scale infection indicator $y(t)$ is then defined as

$$\begin{aligned} y(t) &= \int_0^\infty f_{L_Y}(a)\pi(a,t)dt \\ &= \int_0^\infty f_{L_Y}(a)\pi(0,t-a)dt \\ &= E_{L_Y}[\pi(0,t-L_Y)] \end{aligned} \tag{8}$$

where $E_X[\cdot]$ denotes mathematical expectation with respect to random variable X. Simply stated, the model-scale indicator at time t is the expected

[1]We also considered the generating time density based on the posterior parameter estimates from Park et al's (2020) analysis of generating times, but it did not fit the data quite as well as the generating time based on Li et al (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

value of SARS-CoV-2 incidence L_Y time units into the past. A first-order Taylor approximation yields the approximation

$$y(t) \approx \pi(0, t - E(L_Y)) \tag{9}$$

which suggests that the model-scale indicator value can be approximated by incidence evaluated $E(L_Y)$ time units earlier.

3.1 Example: Hospital Admissions

Hospital admissions have been used as an indicator for the coronavirus outbreak under the presumption that the fraction of new infections that require hospitalization remains constant over time. Define L_H as the time from infection to hospitalization for those infected persons that do require hospital treatment. A review of several published studies by Lewnard et al (2020) estimated that the time from infection until hospitalization averages 13.5 days with 95% probability coverage ranging from 4.8 to 27.9 days. We approximate this finding by employing a gamma distribution with $\alpha = 4.954$ and $\beta = 2.725$ to represent the probability density of L_H, $f_{L_H}(a)$. This distribution also has a mean of 13.5 days with 95% probability coverage ranging from 4.4 to 27.7 days. Similar times from infection to hospitalization are implied by the Centers for Disease Control COVID-19 pandemic planning scenarios (CDC 2020) and also MIDAS (2020).

Conditional upon the transmission model described in equations (4-7), the

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

model-scale hospitalization indicator $h(t)$ is, following equation (8), given by

$$\begin{aligned} h(t) &= \int_0^\infty f_{L_H}(a)\pi(0, t-a)dt \\ &= E_{L_H}[\pi(0, t - L_H)] \\ &\approx \pi(0, t - E(L_H)). \end{aligned} \tag{10}$$

Figure 1 plots both the model-scale SARS-CoV-2 incidence $\pi(0,t)$ and hospitalization indicator $h(t)$ assuming $\lambda(a)$ as defined in equation (3) with $R_0 = 2.38$ as will be estimated subsequently; the gamma distribution for $f_{L_H}(a)$ described above; $s(0) = 1 - 0.0161$ reflecting the initial prevalence of infection as estimated below (with time 0 taken as February 19, 2020); and $\pi(a, 0) = 0.0161/30$ for $0 < a \leq 30$ as explained below. The model-scale hospitalization indicator lags incidence by about two weeks, as one would expect given that $E(L_H) = 13.5$ days by design.

3.2 Example: SARS-CoV-2 RNA in Municipal Sewage Sludge

Peccia et al (2020) reported daily SARS-CoV-2 RNA concentrations based on sampling sludge from a municipal wastewater treatment plant serving the towns of New Haven, East Haven, Hamden, and Woodbridge in the state of Connecticut, USA between March 19 and May 1 of 2020. Virus RNA concentrations in sludge should reflect the amount of virus shed in feces by

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

infected persons in the population served by the treatment plant, resulting in a fecal estimate of community virus RNA concentration. Though virus RNA concentrations in feces degrade exponentially with the time from excretion to sample collection (Hart and Halden, 2020), virus RNA concentrations obtained from sludge sampled daily should be discounted by approximately the same degradation factor, rendering the resulting signal a plausible surrogate tracking community virus RNA concentration over time.

Referring back to the epidemic model, the appropriate measure of virus RNA concentration is the transmission potential $\psi(t)$, as the amount of virus shed in feces should reflect the average infectiousness of the population. However, to use the indicator framework developed above, the age-of-infection transmission rate $\lambda(a)$ must be normalized to the scale of a probability density function. This is easily achieved by defining

$$f_{L_V}(a) = \frac{\lambda(a)}{R_0},\ a > 0 \tag{11}$$

which is immediately recognized as the forward generation time probability density $f(a)$ introduced earlier. This density enables the definition of the model-scale virus RNA indicator $v(t)$ as

$$\begin{aligned} v(t) &= \int_0^\infty f_{L_V}(a)\pi(0, t-a)dt \\ &= E_{L_V}[\pi(0, t - L_V)] \\ &\approx \pi(0, t - E(L_V)). \end{aligned} \tag{12}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

For $\lambda(a)$ as defined in equation (3) using the forward generation time density $f(a)$ corresponding to Li et al (2020), the expected lag $E(L_V)$ is given by 8.9 days, or 4.6 days shorter than the lag from infection to hospitalization. This mean generation time is 3 to 4 days longer than the values implied by CDC (2020) and MIDAS (2020), though it falls within the uncertainty interval estimated by Park et al (2020). Figure 1 reports the model-scale virus RNA indicator $v(t)$ under the same epidemic modeling assumptions described for the hospitalization indicator $h(t)$. Given the transmission model, the timing of both the virus RNA concentration and hospitalization indicators is clear, and provides a clue as to what might be expected when examining the timing of observed hospital admissions and SARS-CoV-2 data in sewage sludge. We turn to such an empirical analysis in the next section.

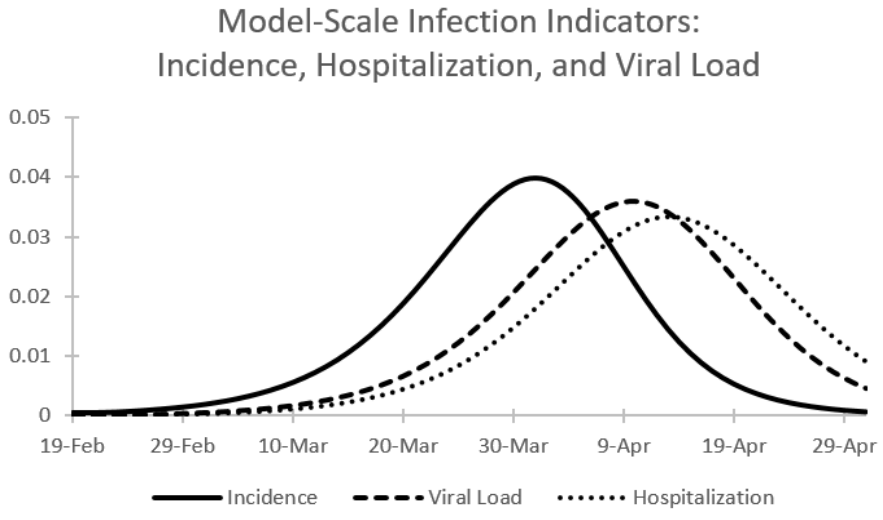

Figure 1

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

4 Scaling Indicators to Transmission: Hospital Admissions and SARS-CoV-2 RNA in Sewage Sludge

Consider a model-scale infection indicator $y(t)$ as earlier described, and let $Y(t)$ be the random variable denoting the real-world scale observable value of this indicator at time t. For examples, corresponding to the model-scale hospitalization indicator $h(t)$, the real-world number of hospital admissions observed on day t is the random variable $H(t)$. Similarly, random variable $V(t)$ denotes the actual concentration of RNA observed in sewage sludge on day t, corresponding with the model-scale virus RNA concentration indicator $v(t)$.

The observable indicator $Y(t)$ is modeled as a random variable with mean proportional to $y(t)$, that is,

$$E(Y(t)) = k_Y y(t),\ t > o \tag{13}$$

for some indicator-specific constant k_Y. We thus scale observable indicators to their model-scale values in expectation. We also allow the indicator variance $\sigma^2_{Y(t)}$ to depend upon $y(t)$. Given the transmission model, we treat $Y(t)$ as conditionally independent of $Y(t')$ for all $t \neq t'$, for correlation in observed values across time would almost entirely be due to the underlying epidemic. The specific probability law presumed for $Y(t)$ given $y(t)$ can differ by infec-

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

tion indicator, as will become clear by example. Given observed indicator values at different points in time and an underlying epidemic model, one can estimate the scaling constants k_Y (and variance parameters if needed) via maximum likelihood or other methods.

The Peccia et al (2020) study of sewage sludge obtained daily COVID-19 admissions data to the Yale New Haven Hospital restricted to residents of the same four Connecticut towns served by the local wastewater treatment plant. The data record the first such admission as occurring on March 14, 2020, 24 days following our February 19 starting date ($t = 0$), and daily admissions data were recorded through May 1, 2020 ($t = 72$) to match the sampling period observed in the sewage sludge study.

Daily hospital admissions data h_t are modeled as realizations of a Poisson random variable $H(t)$ with mean proportional to the model-scale indicator $h(t)$ developed earlier, that is,

$$E[H(t)] = k_H h(t),\, t > 0. \tag{14}$$

The Poisson log likelihood corresponding to the hospital admissions data covering March 14 ($t = 24$) to May 1 ($t = 72$), $\ln \mathcal{L}_H$, is thus given by

$$\ln \mathcal{L}_H = \sum_{t=24}^{72} \left\{ h_t \ln(k_H h(t)) - k_H h(t) \right\}. \tag{15}$$

Also as reported in Peccia et al (2020), sludge samples from the local

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

wastewater treatment plant were tested for SARS-CoV-2 RNA concentrations with two different primers applied to two sample replications daily. These values were adjusted to control for day to day variations in treatment plant flow, sludge solids content, and RNA extraction efficiency (Peccia et al 2020). The data we employ here are v_t, the day t average of these four adjusted values with measurement units 10^5 SARS-CoV-2 RNA copies / ml sludge. Data were collected from March 19 through May 1 ($t = 29, ..., 72$) for a total of 44 daily observations. We model v_t as realizations of a Normal random variable $V(t)$ with mean $E[V(t)] = k_V v(t)$ and variance $\sigma_V^2(t) = c_V v(t)$ to allow for over- or under-dispersion relative to the mean[2]. The Normal log likelihood corresponding to the sludge data, $\ln \mathcal{L}_V$, thus equals

$$\ln \mathcal{L}_V = -\frac{1}{2} \sum_{t=29}^{72} \left\{ \ln(c_V v(t)) + \frac{(v_t - k_V v(t))^2}{c_V v(t)} \right\}. \tag{16}$$

We estimate five parameters from the hospital admissions and sludge data via maximum likelihood, conditional upon the epidemic model (which implies the forward generation lag density $f_{L_V}(a) = f(a)$, and hospital lag density $f_{L_H}(a)$ that was taken from the literature). Three of the parameters estimated are the hospitalization scaling constant k_H, the sludge RNA scaling constant k_V, and the sludge RNA variance scaling constant c_V. The fourth parameter estimated is R_0 which scales the strength of the outbreak and enables direct comparison to SARS-CoV-2 epidemics elsewhere. The final

[2] We also considered a model with constant variance but the model did not fit the data as well.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

parameter estimated is $\pi(0)$, which sets the initial condition of the model via the relation

$$\pi(a, 0) = \pi(0)/30,\ 0 < a \leq 30. \tag{17}$$

This modeling choice reflects the random arrival of imported infections to the area of study in the thirty days preceding the onset of community transmission, in effect determining the placement of the main epidemic wave without changing its shape. A larger value of $\pi(0)$ would pull the epidemic earlier in time, while a smaller value would push the epidemic later. In this way, the hospital admissions and sludge virus RNA concentration data jointly determine the size and the placement of the epidemic wave while impacting the transmission dynamics via the model described in equations (4-7).

Table 1 reports the maximum likelihood estimates and standard errors computed by inverting the Hessian matrix of the log likelihood function (Cox and Hinkley 1974) following maximization of $\ln \mathcal{L}_H + \ln \mathcal{L}_V$, while the fit of the observed data to the scaled model indicators is illustrated in Figures 2 and 3, which plot 95% prediction intervals about the expected indicator values in addition to the data. The hospital admissions data are plotted in Figure 2. Though noisy, the admissions data correspond to the modeled pace of the epidemic, with most values falling within the 95% prediction intervals.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Parameter	Maximum Likelihood Estimate	Standard Error
$\pi(0)$	0.0161	0.0032
k_H	1006.603	56.847
k_V	57.589	4.867
c_V	12.890	2.951
R_0	2.383	0.010

Table 1

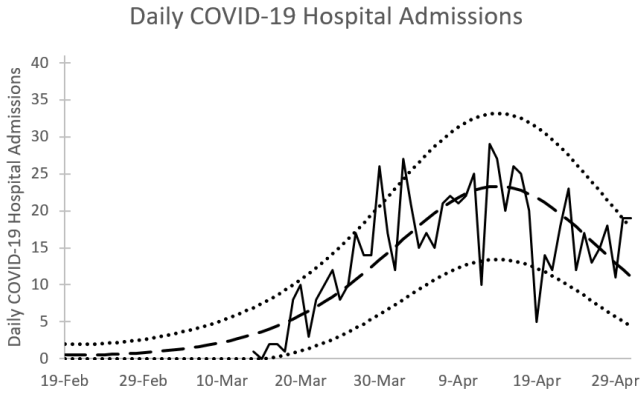

Figure 2

Figure 3 reports the observed and modeled SARS-CoV-2 RNA concentrations (in 10^5 RNA copies / ml sludge) from the sewage study along with conservative 95% prediction intervals. While peak RNA virus concentrations are higher than what would be expected based on the model, the data again match the estimated pace of the epidemic, suggesting that community virus RNA concentration in sewage sludge can indeed be represented by the transmission potential in an epidemic model. The data rise and fall as expected, albeit with much random noise to be sure.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

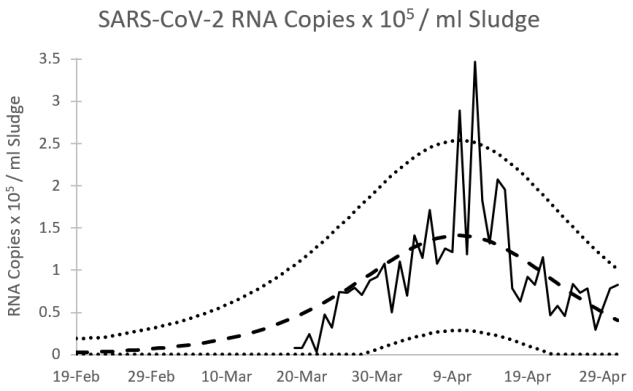

Figure 3

Note that the estimated reproductive number is 2.38 with a 95% confidence interval ranging from 2.19 to 2.58. This places the local SARS-CoV-2 outbreak in New Haven squarely in the middle of reproductive numbers estimated elsewhere (for examples see CDC (2020) and MIDAS (2020)). What is noteworthy is that this reproductive number was estimated from a model linking transmission to hospital admissions and SARS-CoV-2 RNA concentrations measured in sewage sludge. The data, not the model, determined the magnitude of R_0, supporting the plausibility of the hospitalization and generation time lag distributions employed to match the observed data to an underlying transmission model.

Together with the epidemic model, these data answer one of the questions raised by the Peccia et al (2020) study: why did the SARS-CoV-2 RNA signal from the sewage sludge lead hospital admissions by only 4 days, when many were expecting a much earlier signal? The model shows that the natural time lag for virus RNA concentration is governed by the mean forward generation time, estimated at 8.9 days in this model. Given an average 13.5 day lag from

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

infection to hospitalization documented elsewhere, tracking the outbreak by relying on the sewage sludge RNA signal leads similar tracking by hospital admissions by 4.6 days on average.

5 Summary

This paper has focused on modeling lagging epidemic indicators and how they relate to each other. The approach has been to utilize an epidemic model as a basis for scaling indicators like hospital admissions or SARS-CoV-2 RNA observed in sewage sludge. After characterizing how indicators lag incidence, we showed how one could use an epidemic model to simultaneously estimate the placement of an epidemic wave (via estimating the initial condition), the strength of an outbreak (via estimating R_0), and situate lagging indicators appropriately, allowing one to view the data in a more epidemiologically meaningful way. Using data from a recently published study of SARS-CoV-2 RNA concentrations observed in municipal sewage sludge, we showed why the RNA data were only able to shorten the time from infection to signal by 4.6 days relative to hospital admissions. Finally, the RNA and hospitalization data jointly implied an epidemic with R_0 of approximately 2.38, well within the range implied by numerous studies.

Acknowledgements

The authors thank Doug Brackney, Arnau Casanovas-Massana, Nate Grubaugh, Albert Ko, Saad Omer and Dan Weinberger for comments on an earlier draft of this manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

[1] Britton T, Tomba GS (2019) Estimation in emerging epidemics: Biases and remedies. J R Soc Interface 16:20180670. https://doi.org/10.1098/rsif.2018.0670 (accessed July 31, 2020)

[2] CDC (2020). COVID-19 Pandemic Planning Scenarios. United States Centers for Disease Control and Prevention, https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html (accessed July 31, 2020)

[3] Cox DR, Hinkley DV (1974) *Theoretical Statistics*. London: Chapman and Hall.

[4] Champredon D, Dushoff J (2015) Intrinsic and realized generation intervals in infectious-disease transmission. Proc. R. Soc. B 282: 20152026. http://dx.doi.org/10.1098/rspb.2015.2026 (accessed July 31, 2020)

[5] Ferguson N, Laydon D, Gemma N-G et al (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team, March 16, 2020, https://tinyurl.com/tcdy42y (accessed July 31, 2020)

[6] Hart OE, Halden RU. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. Science of The Total Environment (730), 15 August 2020, https://tinyurl.com/yb2dux35 (accessed July 31, 2020).

[7] Heesterbeek JAP, Dietz K (1996) The concept of R_0 in epidemic theory. Statistica Neerlandica 50(1):89-110. https://tinyurl.com/y8m37q9l (accessed July 31, 2020)

[8] Kaplan EH (2020a) Containing 2019-nCoV (Wuhan) coronavirus. Health Care Manag Sci https://doi.org/10.1007/s10729-020-09504-6 (accessed July 31, 2020)

[9] Kaplan EH (2020b) COVID-19 Scratch Models to Support Local Decisions. Manufacturing and Service Operations Management, https://doi.org/10.1287/msom.2020.0891 (accessed July 31, 2020)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[10] Kaplan EH, Forman HP (2020) Logistics of aggressive community screening for coronavirus 2019. JAMA Health Forum (May 4), https://tinyurl.com/y7tmr5x6 (accessed July 31, 2020)

[11] Kissler S, Tedijanto C, Lipsitch M, Grad YH (2020). Social distancing strategies for curbing the COVID-19 epidemic. Harvard School of Public Health, March 2020 http://nrs.harvard.edu/urn-3:HUL.InstRepos:42638988 (accessed July 31, 2020)

[12] Lamont N (2020) Executive Order No. 7J: Protection of pubic health and safety during COVID-19 pandemic and response. https://tinyurl.com/y6u72dmg (accessed July 31, 2020)

[13] Lewnard JA, Liu VX, Jackson ML, PhD5, Schmidt MA, Jewell BL, Flores JP, Jentz C, Northrup GR, Mahmud A, Reingold AL, Petersen M, Jewell NP, Young S, Bellows J (2020) Incidence, clinical outcomes, and transmission dynamics of hospitalized 2019 coronavirus disease among 9,596,321 individuals residing in California and Washington, United States: a prospective cohort study. https://doi.org/10.1101/2020.04.12.20062943 (accessed July 31, 2020)

[14] MIDAS (2020) Online Portal for COVID-19 Modeling Research: COVID-19 parameter estimates. Models of Infectious Disease Agent Study (MIDAS), https://tinyurl.com/qqe5you (accessed July 31, 2020)

[15] Li Q, Guan X, Wu P et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med. https://doi.org/10.1056/NEJMoa2001316 (accessed July 31, 2020)

[16] Park SW, Bolker BM, Champredon D, Earn DJD, Li M, Weitz JS, Grenfell BT, Dushoff J. (2020) Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. J. R. Soc. Interface.1720200144, http://doi.org/10.1098/rsif.2020.0144 (accessed July 31, 2020)

[17] Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, Ko AI, Malik AA, Wang D, Wang M, Warren JL, Weinberger DM, Omer SB (2020). SARS-CoV-2 RNA concentrations in primary municipal sewage sludge

medRxiv preprint doi: https://doi.org/10.1101/2020.06.27.20141739; this version posted August 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

as a leading indicator of COVID-19 outbreak dynamics. https://www.medrxiv.org/content/10.1101/2020.05.19.20105999v2 (accessed July 31, 2020)

[18] Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. B:274, 599–604. https://doi.org/10.1098/rspb.2006.3754 (accessed July 31, 2020)