New Approach to the SIR Inversion Problem: From the 1905-1906 Plague Outbreak in the Isle of Bombay To the 2021-2022 Omicron Surge in New York City

Jeffrey E. Harris MD PhD*†§ Version 2, April 30, 2023

Abstract. We describe a novel approach to recovering the underlying parameters of the SIR dynamic epidemic model from observed data on case incidence or deaths. We formulate a discrete-time approximation to the original continuous-time model and search for the parameter vector that minimizes the standard least squares criterion function. We show that the gradient vector and matrix of second-order derivatives of the criterion function with respect to the parameters adhere to their own systems of difference equations and thus can be exactly calculated iteratively. Applying our new approach, we estimate four-parameter SIR models on two datasets: (1) daily reported cases of COVID-19 during the SARS-CoV-2 Omicron/BA.1 surge of December 2021 - March 2022 in New York City; and (2) weekly deaths from a plague outbreak on the Isle of Bombay during December 1905 - July 1906, originally studied by Kermack and McKendrick in their now-classic 1927 paper. The estimated parameters from the COVID-19 data suggest a duration of persistent infectivity beyond that reported in small-scale clinical studies of mostly symptomatic subjects. The estimated parameters from the plague data suggest that the Bombay outbreak was in fact driven by pneumonic rather than bubonic plague.

Key words. SIR model; inversion problem; Newton-Raphson algorithm; nonlinear least squares; COVID-19; SARS-CoV-2; Omicron variant; bubonic plague; pneumonic plague; structural models; reduced form models; sloppy models; parameter identification; underreporting; heterogeneous mixing

* Professor of Economics Emeritus, Massachusetts Institute of Technology, Cambridge MA USA 02139; and Physician, Eisner Health, Los Angeles CA USA 90015.

† Email: jeffrey@mit.edu

§ See Author Declarations at the end of the manuscript for additional details.

Introduction

Nowadays our sophisticated graphic software can draw attractive plots showing how many people have fallen victim to a highly contagious disease over the course of days, weeks or months. But our graphs alone don't teach us how to reliably determine the underlying risk of transmission from an infected to a susceptible person, or the amount of time that an infected individual remains contagious to others, or what proportion of the population was already infected at the critical point in time when the epidemic wave took off, or how many people remain at risk of infection.

We've just described what mathematicians call the inversion problem [1-4]: how to work backwards from limited data on incident cases or deaths to recover the key parameters underlying our dynamic epidemic models. The problem was born nearly a century ago when Kermack and McKendrick (KM) fit a curve derived from their now-classic model to datapoints of weekly deaths from a plague outbreak on the Isle of Bombay [5]. Since then, scores of investigators have searched for a robust, workable method of estimating the parameters of what has famously come to be known as the SIR (Susceptible-Infected-Removed) model, and the race to find a solution has accelerated with the arrival of the COVID-19 epidemic.

What has made the inversion problem so difficult is that, with some possible exceptions [6-10], the SIR model of coupled differential equations does not admit a closed-form mathematical solution that can be readily used to test the model's predictions against the observed data. That major stumbling block has left us with a motley collection of second-best alternatives.

One idea has been to back out the parameters from the certain salient characteristics of the observed epidemic curve, such as the initial exponential rate of increase of cases [11, 12], the time to reach the peak incidence [13], the rate of decline after the peak [14], and the proportion of the population that is ultimately infected [15-17]. This approach may give us point estimates of the key parameters, but it does not provide any uncertainty ranges around the estimates.

Another idea is to pare down the set of parameters to be identified by making judicious use of prior information on some parameters [18-21], in some cases derived from previous waves of a multi-wave epidemic [22]. Perhaps the most traveled road to a solution has been the use of various parameter search algorithms [18, 23-29] which, when it comes down to it, offer only a marginal improvement over brute force search [30]. Bayesian estimation may be better able to

integrate prior information into our search procedures [31-36], but its computational burden is usually even greater. Last but not least, we can resort to trial and error combined with visual inspection [37].

The present study, we suggest, offers an easily workable solution to the inversion problem. Rather than seeking a closed-form, analytical solution to KM's system of differential equations, we pursue an alternate strategy. First, following the lead of other investigators [35, 38-40], we develop a discrete-time version of their classic, continuous-time SIR model. This step allows us to write their dynamic system in terms of difference equations rather than differential equations. Second, similarly following in others' footsteps [26, 41-44], we define a least squares objective function to test our SIR model's predictions against the observed data.

Third, in what appears to be an innovation, we show that both the gradient vector and Hessian matrix of second-order derivatives of our objective function with respect to the parameters follow their own systems of difference equations. As a result, both the gradient and Hessian can be rapidly and exactly computed by straightforward iteration, an approach that is computationally far superior to numerical approximation [45].

Fourth, once we have calculated the gradient and Hessian, we can use the well-known Newton-Raphson algorithm [46] or Gauss-Newton [47] algorithms to find the global optimum. Fifth, relying on the minimum least squares criterion, we can then calculate the variance-covariance matrix of the parameters and thus determine their confidence intervals [48]. Sixth, our approach permits us to readily determine what parameters are in fact identified when we have only time-series data on new cases or deaths.

We apply our strategy to the estimation of a four-parameter SIR model to two types of datasets. First, we study COVID-19 incidence over a 99-day interval from the December 4, 2021, through March 12, 2022, during the Omicron/BA.1 wave in New York City. Second, we study 31 weeks of data on deaths from the plague from December 17, 1905, through July 15, 1906, during an outbreak in the Isle of Bombay, originally studied by KM [5].

Statistical Methods

Discrete-Time SIR Model

Following the lead of other investigators [4, 35, 38-40], we adopt a discrete-time approach. In our SIR epidemic model, the time axis is marked off in equally spaced intervals $t =$

$0,1,\ldots,T$, where the duration of each interval is sufficiently small as to adequately approximate the classical, continuous-time version [5, 49-52]. At any time t, individuals within this closed population can be in one of three mutually exclusive states: *susceptible* (S), *infected* (I), or *removed* (R). The latter state, which includes both recovered living individuals and decedents, is assumed to be absorbing. To minimize possible complications arising from the non-identifiability of multiple parameters [53-56], we assume a fixed, demographically closed population of size N.

Let S_t, I_t, and R_t denote the respective numbers of individuals in each of the three states at time t. The dynamic path of the epidemic is governed by the following deterministic, coupled difference equations:

$$S_t = S_{t-1} - \beta S_{t-1} I_{t-1}/N \tag{1a}$$

$$I_t = I_{t-1} + \beta S_{t-1} I_{t-1}/N - \gamma I_{t-1} \tag{1b}$$

$$R_t = R_{t-1} + \gamma I_{t-1} \tag{1c}$$

$$S_t + I_t + R_t = N \tag{1d}$$

Equations (1a) through (1d) represent the well-known forward Euler approximation to the underlying continuous-time SIR model of coupled differential equations [57]. Apart from the population size N, this dynamic system has two parameters: β and γ. In equations (1a) and (1b), $\beta > 0$ is an infection transmission parameter, while in equations (1b) and (1c), the parameter $\gamma > 0$ gauges the proportion of infected individuals who transition to the removed state at each time period. The multiplicative term $\beta S_{t-1} I_{t-1}/N$ in equations (1a) and (1b) reflects the law of mass action [58], whereby susceptible individuals become infected in proportion to their frequency of contact with currently infected individuals. All individuals within the population are assumed to mix homogeneously, with no subgroup of individuals mixing preferentially with any other subgroup.

The final equation (1d) reflects the constant size N of the population and is consistent with equations (1a) through (1c). Strictly speaking, one should adjust the size N of the mixing population in equations (1a) and (1b) to take account of removals by death. Unless the overall death rate is substantial, this adjustment is usually ignored in model implementations.

To further simplify matters, we assume that initially all individuals are either susceptible or infected. Specifically, at $t = 0$, we have:

$$S_0 = (1 - i_0)N \tag{2a}$$

$$I_0 = i_0 N \tag{2b}$$

$$R_0 = 0 \tag{2c}$$

In equations (2a) and (2b), the additional parameter i_0 represents the proportion of the entire population of size N that is initially infected. We address the more general case where $R_0 > 0$ below.

In the dynamic system (1), the mean duration of infection is time-invariant and equal to $1/\gamma$. Given the initial conditions (2), the system will result in an epidemic wave so long as $(1 - i_0)\,\beta/\gamma > 1$, a well-known result known as the epidemic threshold theorem [24, 51, 59, 60]. Assuming that $1 - i_0 \approx 1$, most authors write this epidemic threshold condition as $\mathcal{R}_0 > 1$, where $\mathcal{R}_0 = \beta/\gamma$ is the basic reproduction number [24, 52, 61, 62].

Parameter Estimation Problem

We do not have direct observations on the underlying *state variables* S_t, I_t, and R_t. If we had such data, our inversion problem would border on trivial [39, 63, 64]. Instead, we observe only the reported counts of new infections at various intervals. For now, we assume that new infection counts y_t are observed at each discrete time t. Below we address the more general case where such counts are observed less frequently. The counts y_t represent observations on the *output variables* X_t, $t = 1, \ldots, T$, which from (1) correspond to:

$$X_t = \beta S_{t-1} I_{t-1}/N \tag{3}$$

It will be helpful to define the parameter $\alpha = 1 - \gamma$ as the proportion of infected individuals who remain infected from one time period to the next, so that the probability an infected individual remains infectious after τ time periods is α^τ. Given the definition of the output variable X_t in (3), our dynamic system (1) can be redefined as:

$$S_t = S_{t-1} - X_t \tag{4a}$$

$$I_t = X_t + \alpha I_{t-1} \tag{4b}$$

So long as the state variables adhere to the condition that that $S_t + I_t + R_t = N$, an explicit difference equation for R_t is unnecessary.

We can now characterize our parameter estimation problem. Given observations y_t on the output variables X_t, we want to estimate the unknown parameters β, α, i_0, and N. Our stumbling block is that we cannot express X_t as a closed-form function of these parameters. We know only that the output variables X_t adhere to the dynamic system defined by (2), (3) and (4), which in turn depends on the parameters β, α, i_0, and N.

Solution Strategy: Nonlinear Least Squares

To proceed, we need some additional notation. Let $\boldsymbol{y} = (y_1, \dots, y_T)'$ and $\boldsymbol{X} = (X_1, \dots, X_T)'$, respectively, denote column vectors of the observed incidence data and the corresponding output variables at each time t, where we use boldface symbols denote vectors or matrices. Let $\boldsymbol{\theta} = (\beta, \alpha, i_0)'$ denote the column vector of the unknown parameters excluding the population size N. For now, we condition on N, but as we discuss below, this additional parameter can be separately identified once $\boldsymbol{\theta}$ has been estimated. Let $\boldsymbol{X}(\boldsymbol{\theta})$ represent the functional dependence of the output variables on the parameters, suppressing for now that the fact that $\boldsymbol{X}$ also depends on N. Finally, let the operator $\boldsymbol{D}$ denote the gradient of partial derivatives with respect to the parameters. For example, the column vector $\boldsymbol{DX_t} = \left(\frac{\partial X_t}{\partial \beta}, \frac{\partial X_t}{\partial \alpha}, \frac{\partial X_t}{\partial i_0}\right)'$ represents the gradient of partial derivatives of the output variable X_t at time t.

We want to optimize some objective function $V\big(\boldsymbol{y}, \boldsymbol{X}(\boldsymbol{\theta})\big)$ with respect to $\boldsymbol{\theta}$. While $\boldsymbol{X}(\boldsymbol{\theta})$ has no closed-form expression, the time-specific gradient vectors $\boldsymbol{DX_t}$ of partial derivatives of X_t with respect to the elements of $\boldsymbol{\theta}$ follow a system of difference equations and can thus be readily calculated iteratively. The same goes for the Hessian matrices $\boldsymbol{D^2X_t}$ of second-order partial derivates with respect to the elements of $\boldsymbol{\theta}$. Once we have computed the gradient vectors $\boldsymbol{DX_t}$ and Hessian matrices $\boldsymbol{D^2X_t}$, we can use the Newton-Raphson algorithm [46] or, under more restrictive conditions, the Gauss-Newton [47] algorithm to optimize V.

To see how this solution strategy works, we specifically consider the nonlinear least squares optimization criterion $V\big(\boldsymbol{y}, \boldsymbol{X}(\boldsymbol{\theta})\big) = \big(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\theta})\big)'\big(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\theta})\big)$, which can be written in summation notation as:

$$V = \sum_{t=1}^{T}\big(y_t - X_t(\boldsymbol{\theta})\big)^2 \qquad (5)$$

This criterion has been widely used in attempts to fit the SIR model to incidence data [26, 41-44]. It is well known that minimizing V is equivalent to computing the maximum likelihood estimate of $\boldsymbol{\theta}$ under the assumption that the observations y_t are independently normally distributed $N(X_t(\boldsymbol{\theta}), \sigma^2)$ with respective means $X_t(\boldsymbol{\theta})$ and homoscedastic variance σ^2. In Appendix A, we adopt a more general maximum likelihood framework and specifically consider an alternative Poisson distribution.

Calculating the Gradient of the Least Squares Optimization Criterion V

The gradient of the least squares criterion V in equation (5) with respect to the parameter vector $\boldsymbol{\theta}$ is:

$$\boldsymbol{DV} = -2\sum_{t=1}^{T}(y_t - X_t)\,\boldsymbol{DX_t} \tag{6}$$

From (6), we learn that the gradient $\boldsymbol{DV}$ depends on the gradients $\boldsymbol{DX_t}$. Our next step is to derive specific expressions for the latter gradient terms. Taking the derivative of (3) with respect to $\boldsymbol{\theta}$, we get:

$$\boldsymbol{DX_t} = \left(S_{t-1}\frac{I_{t-1}}{N}\right)\boldsymbol{D\beta} + \left(\frac{\beta}{N}\right)I_{t-1}\boldsymbol{DS_{t-1}} + \left(\frac{\beta}{N}\right)S_{t-1}\boldsymbol{DI_{t-1}} \tag{7},$$

where the notation $\boldsymbol{D\beta}$ represents the unit column vector $\left(\frac{\partial\beta}{\partial\beta}, \frac{\partial\beta}{\partial\alpha}, \frac{\partial\beta}{\partial i_0}\right)' = (1,0,0)'$. Taking the derivative of (1a), we get:

$$\boldsymbol{DS_t} = \boldsymbol{DS_{t-1}}\left(1 - \left(\frac{\beta}{N}\right)I_{t-1}\right) - S_{t-1}\frac{I_{t-1}}{N}\boldsymbol{D\beta} - S_{t-1}\left(\frac{\beta}{N}\right)\boldsymbol{DI_{t-1}} \tag{8}$$

From (1b) and (3), we can write $I_t = \alpha I_{t-1} + X_t = \alpha I_{t-1} + \beta S_{t-1}I_{t-1}/N = \left(\alpha + \frac{\beta}{N}S_{t-1}\right)I_{t-1}$. Taking the derivative of this expression gives:

$$\boldsymbol{DI_t} = \left(\alpha + \frac{\beta}{N}S_{t-1}\right)\boldsymbol{DI_{t-1}} + I_{t-1}\left(\boldsymbol{D\alpha} + \frac{S_{t-1}}{N}\boldsymbol{D\beta}\right) + I_{t-1}\left(\frac{\beta}{N}\right)\boldsymbol{DS_{t-1}} \tag{9},$$

where $\boldsymbol{D\alpha}$ similarly represents the unit vector $(0,1,0)'$.

We next need to compute the initial values of the gradients $\boldsymbol{DS_0}$ and $\boldsymbol{DI_0}$. Since $I_0 = i_0 N$ and $S_0 = (1 - i_0)N$, we can write the gradients $\boldsymbol{DS_0} = -N\boldsymbol{Di_0}$ and $\boldsymbol{DI_0} = N\boldsymbol{Di_0}$, where the gradient $\boldsymbol{Di_0}$ simplifies to $(0,0,1)'$. Given these initial gradient values, we can use (8) and (9) to iteratively compute the vectors $\boldsymbol{DS_t}$ and $\boldsymbol{DI_t}$ for all $t = 1, \ldots, T$. Once these gradients have been

computed, we can then apply (7) to iteratively compute the corresponding gradients $\boldsymbol{DX_t}$. Those quantities in turn yield the gradient $\boldsymbol{DV}$ of our objective function through equation (6).

Calculating the Hessian Matrix

Let $\boldsymbol{DX}$ denote the $T \times 3$ matrix whose t-th row is the vector $\boldsymbol{DX_t}'$. Taking the derivative of the gradient in (6), we obtain:

$$\boldsymbol{D^2V} = 2\,\boldsymbol{DX'DX} - 2\sum_{t=1}^{T}(y_t - X_t)\boldsymbol{D^2X_t} \tag{10}$$

From (10), we similarly learn that the Hessian $\boldsymbol{D^2V}$ depends on the gradients $\boldsymbol{DX_t}$, which we have already derived, as well the Hessian matrices $\boldsymbol{D^2X_t}$.

To develop the corresponding iterative formulas for the Hessian matrices $\boldsymbol{D^2X_t}$, we need some additional notation. For two column vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ of dimension $L \times 1$, we define the outer product $\boldsymbol{A} \cdot \boldsymbol{B}$ as the $L \times L$ matrix $\boldsymbol{AB'}$ formed by the cross-products of the elements of $\boldsymbol{A}$ and $\boldsymbol{B}$. (Note that $\boldsymbol{B} \cdot \boldsymbol{A}$ is the transpose of $\boldsymbol{A} \cdot \boldsymbol{B}$.) We further define $\boldsymbol{A} \odot \boldsymbol{B} = (\boldsymbol{A} \cdot \boldsymbol{B}) + (\boldsymbol{B} \cdot \boldsymbol{A})$ as the symmetric matrix formed by the sum of the outer product and its transpose. In the three-dimensional case, $\boldsymbol{A} = (a_1, a_2, a_3)'$, and $\boldsymbol{B} = (b_1, b_2, b_3)'$, we have:

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} a_1b_1 + b_1a_1 & a_1b_2 + b_1a_2 & a_1b_3 + b_1a_3 \\ a_2b_1 + b_2a_1 & a_2b_2 + b_2a_2 & a_2b_3 + b_3a_2 \\ a_3b_1 + b_3a_1 & a_3b_2 + b_3a_2 & a_3b_3 + b_3a_3 \end{bmatrix} \tag{11}$$

Now taking the derivative of $\boldsymbol{DX_t}$ as displayed in (7), we get:

$$\boldsymbol{D^2X_t} = \left(\frac{1}{N}\right)(S_{t-1}\boldsymbol{DI_{t-1}} + I_{t-1}\boldsymbol{DS_{t-1}}) \odot (\boldsymbol{D\beta}) +$$

$$\left(\frac{\beta}{N}\right)\left((\boldsymbol{DI_{t-1}}) \odot (\boldsymbol{DS_{t-1}}) + S_{t-1}\boldsymbol{D^2I_{t-1}} + I_{t-1}\boldsymbol{D^2S_{t-1}}\right) \tag{12}$$

The Hessian matrices $\boldsymbol{D^2X_t}$ depend in turn on the Hessian matrices $\boldsymbol{D^2S_t}$ and $\boldsymbol{D^2I_t}$. The corresponding expressions for these Hessian matrices are:

$$\boldsymbol{D^2S_t} = \left(1 - \left(\frac{\beta}{N}\right)I_{t-1}\right)\boldsymbol{D^2S_{t-1}} - \left(\frac{\beta}{N}\right)(\boldsymbol{DI_{t-1}}) \odot (\boldsymbol{DS_{t-1}}) - \left(\left(\frac{\beta}{N}\right)S_{t-1}\right)\boldsymbol{D^2I_{t-1}}$$

$$-\left(\frac{1}{N}\right)(S_{t-1}\boldsymbol{DI_{t-1}} + I_{t-1}\boldsymbol{DS_{t-1}}) \odot (\boldsymbol{D\beta}) \tag{13}$$

$$\boldsymbol{D^2I_t} = \left(\alpha + \frac{\beta}{N}S_{t-1}\right)\boldsymbol{D^2I_{t-1}} + \left(\frac{\beta}{N}\right)(\boldsymbol{DI_{t-1}}) \odot (\boldsymbol{DS_{t-1}}) + \left(\left(\frac{\beta}{N}\right)I_{t-1}\right)\boldsymbol{D^2S_{t-1}}$$

$$+\left(\frac{1}{N}\right)(S_{t-1}\boldsymbol{DI_{t-1}} + I_{t-1}\boldsymbol{DS_{t-1}}) \odot (\boldsymbol{D\beta}) + \boldsymbol{DI_{t-1}} \odot (\boldsymbol{D\alpha}) \tag{14}$$

To complete our calculations, we note that the initial values $\boldsymbol{D^2 S_0}$ and $\boldsymbol{D^2 I_0}$ are simply null matrices.

Taken together, equations (10), (12), (13) and (14), along with the initial null values of $\boldsymbol{D^2 S_0}$ and $\boldsymbol{D^2 I_0}$, permit us to iteratively compute the Hessian $\boldsymbol{D^2 V}$, just as we outlined for the gradient $\boldsymbol{DV}$ above.

Multiple Local Optima; Highly Correlated Parameter Estimates

In the Newton-Raphson [46] approach to optimization, which we employed here, the current value of the parameter vector $\boldsymbol{\theta}^{(k)}$ is repeatedly mapped into an updated value $\boldsymbol{\theta}^{(k+1)}$ according to the well-known rule:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - q\ (\boldsymbol{D^2 V})^{-1}\ \boldsymbol{DV} \qquad (15)$$

In equation (15), the gradient $\boldsymbol{DV}$, as defined in (6), and the Hessian $\boldsymbol{D^2 V}$, as defined in (10), are both computed at the current value of the parameter vector $\boldsymbol{\theta}^{(k)}$. The step size $0 < q \leq 1$ is under control of the programmer. This approach works flawlessly when the objective function V is globally convex, for example, when each output variable $\boldsymbol{X_t}$ is a linear function of $\boldsymbol{\theta}$. But that's hardly the situation in the SIR inversion context.

To begin with, there is good reason to suspect that V is not globally convex and may have multiple local optima. While the classical SIR model predicts a single-peaked wave of incident cases X_t so long as $\mathcal{R}_0 > 1$, the observed data y_t often display multiple peaks over time. (A good example is the two-peak plot of the 2001 Dengue fever outbreak in Havana [60].) In that case, the search algorithm embodied in (15) may easily end up at a local rather than a global optimum of V as it gets trapped in a region of the parameter space that fits one of the multiple peaks.

What's more, there will be regions of the parameter space where the Hessian $\boldsymbol{D^2 V}$ is not positive definite because the second term in equation (11) (that is, $-2\sum_{t=1}^{T}(y_t - X_t)\boldsymbol{D^2 X_t}$) will not necessary be a positive definite matrix. In that case, the updated parameter rule (15) may not result in a decrease in the objective function V as the algorithm veers away from the optimum.

These difficulties may be partly addressed by the Gauss-Newton algorithm [47], which in the current context entails the following parameter updating rule:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - q\ (2\ \boldsymbol{DX}'\boldsymbol{DX})^{-1}\ \boldsymbol{DV} \qquad (16)$$

The idea here is that matrix 2 $\boldsymbol{DX'DX}$, which corresponds to the first term in our expression for the Hessian $\boldsymbol{D^2V}$ in (10), is always positive definite, so that the rule (16) will consistently result in a decrease in V. However, the matrix $\boldsymbol{DX'DX}$ may be ill-conditioned, with eigenvalues spanning multiple orders of magnitude, so that the steps $\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}$ jump around uncontrollably and fail to converge to zero. While various modifications such the Levenberg-Marquardt algorithm have been proposed to recondition the matrix $\boldsymbol{DX'DX}$ [65], in the final analysis our search procedure may end up being exquisitely sensitive to the choice of the initial parameter vector $\boldsymbol{\theta}^{(0)}$.

Even when our objective function V has an interior global minimum, where it is at least locally convex in the neighborhood surrounding the optimum, we may still encounter an inverted ridge or ravine where the parameters are highly correlated. There may indeed be a unique value $\boldsymbol{\theta}^*$ at the very bottom of the ravine that achieves the global minimum V^*. But along the floor of the ravine, there is a one-dimensional curve containing $\boldsymbol{\theta}^*$ where V is almost equal to V^*. Strictly speaking, the parameters are structurally identified, but they are not *practically* identified [66].

To confront these problems of multiple equilibria and correlated parameters in the applications below, we took care to construct and inspect maps of the least-squares criterion V and the gradient $\boldsymbol{DV}$ as functions of the parameters. We also examined the path of successive updates $\boldsymbol{\theta}^{(k)}$ in parameter space generated by our Newton-Raphson search algorithm.

Confidence Intervals; Reparametrizing

Our reliance on the nonlinear least squares criterion V in equation (5) permits us to estimate confidence intervals for the estimated parameters [48]. Conditional upon N, the variance-covariance matrix of the parameters $\boldsymbol{\theta} = (\beta, \alpha, i_0)$ is $\boldsymbol{C} = s^2(\boldsymbol{DX'DX})^{-1}$, where $s^2 = V/T$, and where $\boldsymbol{DX}$ and V are evaluated at the optimum. The standard errors are the square roots of the diagonal elements of $\boldsymbol{C}$. The symmetric 95 percent confidence intervals, based upon the assumption of an asymptotic normal distribution, can be evaluated as ± 1.96 standard errors about the estimates. We can then use the Delta method [67] to compute the standard errors and confidence intervals around nonlinear functions of the parameters, such as the basic reproduction number $\mathcal{R}_0 = \beta/\gamma = \beta/(1-\alpha)$ and the mean duration of infection $1/\gamma = 1/(1-\alpha)$.

In some applications where the data y_t exhibit substantial variability and the resulting value of s^2 is large, the estimated symmetric confidence interval surrounding α may exceed the

allowable range of $0 \leq \alpha < 1$. In such cases, we can reparametrize, instead specifying $\boldsymbol{\theta} = (\beta, \omega, i_0)$ with $\alpha(\omega) = 1/(1 + e^{-\omega})$, where the substitute parameter ω is unbounded. Our expression for $\boldsymbol{DI_t}$ in equation (9) will remain valid, but with the modification that $\boldsymbol{D\alpha} = \left(0, \frac{d\alpha}{d\omega}, 0\right) = (0, \alpha(1-\alpha), 0)$. Our expression for $\boldsymbol{D^2 I_t}$ in equation (14) will become:

$$\boldsymbol{D^2 I_t} = \left(\alpha + \frac{\beta}{N} S_{t-1}\right) \boldsymbol{D^2 I_{t-1}} + \left(\frac{\beta}{N}\right) (\boldsymbol{DI_{t-1}}) \odot (\boldsymbol{DS_{t-1}}) + \left(\left(\frac{\beta}{N}\right) I_{t-1}\right) \boldsymbol{D^2 S_{t-1}} +$$

$$\left(\frac{1}{N}\right) (S_{t-1} \boldsymbol{DI_{t-1}} + I_{t-1} \boldsymbol{DS_{t-1}}) \odot (\boldsymbol{D\beta}) + \boldsymbol{DI_{t-1}} \odot (\boldsymbol{D\alpha}) + I_{t-1} \boldsymbol{D^2 \alpha} \qquad (17)$$

The additional term on the right-hand side of equation (17), which does not appear in (14), contains the matrix $\boldsymbol{D^2 \alpha} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{d^2\alpha}{d\omega^2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$, where $\frac{d^2\alpha}{d\omega^2} = \alpha(1-\alpha)(1-2\alpha)$.

Recovering the Population Size Parameter N

While $\boldsymbol{X}(\boldsymbol{\theta})$ has no closed-form expression in terms of $\boldsymbol{\theta}$, it turns out that each element X_t is a linear function of the population size parameter N. That is, X_t can be written in the form $\varphi_t(\boldsymbol{\theta})N$, where $\varphi_t(\boldsymbol{\theta})$ is a function of the remaining parameters $\boldsymbol{\theta}$.

To prove this result, we only need to show that the state variables S_t and I_t are themselves proportional to N. Thus, if for all times $t = 1, \ldots, T$, we can show that $S_t = f_t(\boldsymbol{\theta})N$ and $I_t = g_t(\boldsymbol{\theta})N$ for some functions $f_t(\boldsymbol{\theta})$ and $g_t(\boldsymbol{\theta})$, then from equation (3), we would have $X_t = \beta S_{t-1} I_{t-1}/N = \beta f_{t-1}(\boldsymbol{\theta}) g_{t-1}(\boldsymbol{\theta}) N$. Thus, $\varphi_t(\boldsymbol{\theta}) = \beta f_{t-1}(\boldsymbol{\theta}) g_{t-1}(\boldsymbol{\theta})$.

We can prove that S_t and I_t are proportional to N by mathematical induction. From equations (2a) and (2b), respectively, we known that $S_0 = (1 - i_0)N$ and $I_0 = i_0 N$. So, the proposition is true for $t = 0$. Now suppose that $S_t = f_t(\boldsymbol{\theta})N$ and $I_t = g_t(\boldsymbol{\theta})N$, at time t. We claim that $S_{t+1} = f_{t+1}(\boldsymbol{\theta})N$ and $I_{t+1} = g_{t+1}(\boldsymbol{\theta})N$ necessarily hold for some functions $f_{t+1}(\boldsymbol{\theta})$ and $g_{t+1}(\boldsymbol{\theta})$. From (1a), we can write $S_{t+1} = S_t \left(1 - \frac{\beta}{N} I_t\right)$. We have $S_{t+1} = S_t \left(1 - \frac{\beta}{N} I_t\right) = f_t(\theta) N \left(1 - \frac{\beta}{N} g_t(\theta) N\right) = \big(f_t(\theta) - \beta g_t(\theta)\big) N$, and so $f_{t+1}(\theta) = f_t(\theta) - \beta g_t(\theta)$. Similarly, $I_{t+1} = I_t \left(\alpha + \frac{\beta}{N} S_t\right) = g_t(\theta) N \left(\alpha + \frac{\beta}{N} f_t(\theta) N\right) = g_t(\theta)\big(\alpha + \beta f_t(\theta)\big) N$, and so $g_{t+1}(\theta) = g_t(\theta)\big(\alpha + \beta f_t(\theta)\big)$. ■

The fact that each output variable X_t is proportional to N permits us to employ an iterative procedure resembling the expectation-maximization (or EM) algorithm [68] to recover

an estimate of that parameter. Let's say that we have a parameter estimate $N^{(n)}$ at iteration n of the algorithm. We employ the previously described procedure to estimate $\boldsymbol{\theta}^{(n)}$ conditional upon $N^{(n)}$. That corresponds to the maximization (or M) step. (From the coding standpoint, when apply the Newton-Raphson parameter search rule described in (15) to estimate $\boldsymbol{\theta}^{(n)}$, we're looping on k within a loop on n.) Given both $N^{(n)}$ and $\boldsymbol{\theta}^{(n)}$, we have output-variable estimates $X_t^{(n)} = \varphi_t\left(\boldsymbol{\theta}^{(n)}\right)N^{(n)}$ for all t. We now compute:

$$\kappa^{(n+1)} = \boldsymbol{y}'\boldsymbol{X}^{(n)}/\boldsymbol{X}^{(n)\prime}\boldsymbol{X}^{(n)} = \sum_{t=1}^{T} y_t X_t^{(n)} / \sum_{t=1}^{T}\left(X_t^{(n)}\right)^2 \tag{18}$$

That is, we compute an adjustment factor $\kappa^{(n+1)}$ by regressing $\boldsymbol{y}$ on $\boldsymbol{X}^{(n)}$. For the expectation (or E) step, we update the population size as $N^{(n+1)} = \kappa^{(n+1)}N^{(n)}$. This iterative procedure will work for any optimization criterion V and not just for the least squares criterion described above.

Non-Identifiability of the Initially Resistant Population

We have thus far assumed in equation (2c) that no one in the population was initially resistant to infection, that is, $R_0 = 0$. Now let's assume more generally that a non-negative fraction $1 > r_0 \geq 0$ of the population is already resistant at $t = 0$. We thus replace the restricted initial conditions (2) with the following more general initial conditions:

$$S_0 = (1 - i_0 - r_0)N \tag{19a}$$

$$I_0 = i_0 N \tag{19b}$$

$$R_0 = r_0 N \tag{19c}$$

So long as the R_0 initially resistant individuals mix homogeneously with those in the susceptible and infected states, equation (3) governing the output variables X_t as well as the equations of motion (4) for the state variables S_t and I_t remain unchanged.

Apart from our original parameter vector $\boldsymbol{\theta} = (\beta, \alpha, i_0)$ and the population size N, it appears that we now have an additional parameter r_0 to be estimated. It turns out, however, that r_0 cannot be identified in our discrete SIR model from the observed data $\boldsymbol{y}$ alone [54].

To see why, let's suppose that we estimate our more general model conditional upon some fixed value of r_0. We denote the resulting minimum least squares estimates as $\boldsymbol{\theta}^*(r_0)$ and $N^*(r_0)$ to show their dependence on the value of r_0 chosen. We denote the corresponding individual components of $\boldsymbol{\theta}^*(r_0)$ as $\beta^*(r_0)$, $\alpha^*(r_0)$, and $i_0^*(r_0)$. Our original parameter estimates

conditional upon $r_0 = 0$ thus correspond to $\beta^*(0)$, $\alpha^*(0)$, $i_0^*(0)$, and $N^*(0)$. Then the following conditions hold for all $1 > r_0 \geq 0$:

$$\beta^*(r_0) = \beta^*(0)/(1 - r_0) \tag{20a}$$

$$\alpha^*(r_0) = \alpha^*(0) \tag{20b}$$

$$i_0^*(r_0) = i_0^*(0)(1 - r_0) \tag{20c}$$

$$N^*(r_0) = N^*(0)/(1 - r_0) \tag{20d}$$

What's more, the paths of the output variable X_t and the state variables S_t and I_t at the optimum values of the parameters, but not the state variable R_t, will be independent of r_0. Accordingly, the optimum value of V will likewise be independent of r_0.

We again rely upon mathematical induction to prove that the state variables S_t and I_t at the optimum are independent of r_0. First consider $t = 0$. For any value of r_0, we have from (19a) that $S_0 = (1 - i_0^*(r_0) - r_0)N^*(r_0)$. Applying (20c) and (20d), we get $S_0 = \big(1 - i_0^*(0)\big)N^*(0)$, which does not depend on r_0. Similarly for any value of r_0, we have from (19b) that $I_0 = i_0^*(r_0)N^*(r_0)$. Again applying (20c) and (20d), we get $I_0 = i_0^*(0)N^*(0)$, which is likewise independent of r_0. Now assume that S_t and I_t are independent of r_0. We show that S_{t+1} and I_{t+1} must also be independent of r_0. From (1a), we have $S_{t+1} = S_t(1 - \beta^*(r_0)I_t/N^*(r_0))$. Applying (19a) and (19d), this expression resolves to $S_{t+1} = S_t(1 - \beta^*(0)I_t/N^*(0))$. From (1b) and (3), we have $I_{t+1} = I_t(\alpha + \beta^*(r_0)S_t/N^*(r_0))$. Again applying (20a) and (20d), this expression similarly resolves to $I_{t+1} = I_t(\alpha + \beta^*(0)S_t/N^*(0))$. ■

Interpreting the Estimated Population Size N: Underreporting and Incomplete Mixing

While the additional parameter r_0 is not identifiable from the data $\boldsymbol{y}$ alone, we might still be able to identify it from other data. Suppose, for example, that we had sharp prior information on the basic reproduction number $\mathcal{R}_0 = \beta/\gamma = \beta/(1 - \alpha)$. Let's denote this prior estimate $\bar{\mathcal{R}}_0$. We first compute the basic reproductive number $\mathcal{R}_0^*(0) = \beta^*(0)/\big(1 - \alpha^*(0)\big)$ implied by our parameter estimates conditional upon $r_0 = 0$. So long as $\bar{\mathcal{R}}_0 \geq \mathcal{R}_0^*(0)$, we can then use our prior information to conclude that $r_0 = 1 - \mathcal{R}_0^*(0)/\bar{\mathcal{R}}_0$.

Why couldn't we similarly take advantage of prior information on the population size to ascertain the initial proportion r_0 of resistant individuals? Suppose we knew from census data that the population contained $\bar{N}$ individuals. We would first estimate $N^*(0)$ conditional upon

$r_0 = 0$, and then, so long as $\overline{N} \geq N^*(0)$, we would have the estimate $r_0 = 1 - N^*(0)/\overline{N}$. Unfortunately, reliance solely on census data for prior estimates of population size N is complicated by two other phenomena: underreporting and incomplete mixing.

In many contexts, particularly in recent applications of the SIR and related compartmental models to COVID-19 incidence, it has been widely recognized that a significant number of incident cases may have gone unreported [69]. In the absence of reliable information on the temporal pattern of such underreporting, the most parsimonious approach to this phenomenon has been to assume a *case identification ratio* equal to a constant $p < 1$ [70-72]. In that case, our model would need to be modified to accommodate the reality that our reported incidence data y_t are in fact estimates of pX_t rather than X_t. We've already learned, however, that $X_t = \varphi_t(\boldsymbol{\theta})N$, where $\varphi_t(\boldsymbol{\theta})$ is a function of the remaining parameters $\boldsymbol{\theta}$, and this is the case even in our more general model admitting $r_0 > 0$. Accordingly, our reported incidence data y_t are really estimates of $pX_t = \varphi_t(\boldsymbol{\theta})(pN)$, and the estimated population size derived from our model is really an estimate of pN.

We thus have a knotty problem of confounding. We can estimate population size $N^*(0)$ assuming that there are no initially resistant individuals and no underreporting. If we had census data $\overline{N}$, we could account for both phenomena, writing $N^*(0)/\overline{N} = p(1 - r_0)$. Without more information, we cannot separately identify p and r_0.

We would ordinarily interpret the parameter N to gauge the size of the population at risk for contagion. This population would consist of all individuals who homogeneously mix with each other in accordance with the law of mass action embodied in equations (1) and (3). Many investigators, however, have properly recognized that the underlying assumption of homogeneous mixing may not apply to the entire population [19, 73-78].

Let's say we're analyzing an outbreak on a college campus with known student enrollment $\overline{N}$. When we run our model with $r_0 = 0$, we obtain an estimate $N^*(0)$ that is, say, only about 10 percent of $\overline{N}$. This finding does not necessarily imply that 90 percent of the student body was resistant or that only one in ten cases was reported. Instead, it may mean that only small fraction of the student body was directly involved in the mixing process that generated the outbreak. A good example would be the COVID-19 outbreak at the campus of the University of Wisconsin-Madison in September 2020, where total student enrollment was $\overline{N} = 44{,}640$, but

where the large fraction of cases was concentrated in two on-campus student residence halls with a combined population of about 2,932 [79].

Any comparison between the estimated population size $N^*(0)$ and the census statistic $\bar{N}$ thus entails three confounded interpretations. Some cases have gone unreported. Some individuals may be initially resistant. And other individuals may be susceptible but remain outside the core population through which the infectious agent has spread.

Adapting the Model to Data on Removals

Our analysis has thus far assumed that each data point y_t is an observation on the corresponding output variable X_t capturing the number of individuals who transitioned from the *susceptible* to the *infected* state at time t. Alternatively, we can construct an analogous model where the data point y_t is instead an observation on a distinct output variable Z_t capturing the number of individuals who transitioned from the *infected* to the *removed* state at time t. This approach would be important if we had data on the number of deaths during an epidemic of a disease with a near 100 percent fatality rate, as KM observed in the case of 1905-1906 plague outbreak on the Isle of Bombay [5].

In this alternative model, the output variable Z_t is:

$$Z_t = R_t - R_{t-1} = (1-\alpha)I_t \tag{21}$$

Continuing with the least squares framework, the objective function is now:

$$V = \sum_{t=1}^{T}\left(y_t - Z_t(\boldsymbol{\theta})\right)^2 \tag{22}$$

To minimize this objective function employing the same strategy above, we will need to compute the gradient vector $\boldsymbol{DZ_t}$ and the Hessian matrix $\boldsymbol{D^2Z_t}$. From (25), we immediately have:

$$\boldsymbol{DZ_t} = (1-\alpha)\boldsymbol{DI_t} - I_t\boldsymbol{D\alpha} \tag{23}$$

The corresponding Hessian matrix is:

$$\boldsymbol{D^2Z_t} = (1-\alpha)\boldsymbol{D^2I_t} - \boldsymbol{DI_t}\odot\boldsymbol{D\alpha} \tag{24}$$

To compute these two expressions, we will need the gradient $\boldsymbol{DI_t}$ and Hessian $\boldsymbol{D^2I_t}$, but these respective quantities were already computed in equations (9) and (14) above. To complete our estimation procedure, we note that the output variables are similarly linear functions of the population size parameter N, so that the EM-type algorithm described above likewise applies.

Working with Aggregate Data

In some applications, we may have only aggregate data reported over a coarse time scale, rather than the finely divided time axis assumed so far. For concreteness, let's assume that the underlying SIR epidemic model is valid when the time axis $t = 0,1, \dots, T$ is marked off in days, but we only have data y_m on the cumulative number of deaths during each 7-day week, indexed by $m = 1, \dots, M$. Days t are mapped into weeks m by the relation $m = h(t) = \left\| \frac{t-1}{7} \right\| + 1$, where the floor operator $\|x\|$ maps into the largest integer j such that $j \leq x$. We further assume that $T = 7M$, so that week M, which corresponds to the last week, ends on day T.

Now define the $M \times T$ aggregation matrix $\boldsymbol{W}$ with element $w_{mt} = 1$ if and only if $m = h(t)$. Otherwise, $w_{mt} = 0$. While our finely disaggregated model generates daily values Z_t of the output variable from the underlying parameters $\boldsymbol{\theta}$, our data y_m represent observations on the aggregated values $\boldsymbol{w_m Z} = \sum_{t=1}^{T} w_{mt} Z_t$, where $\boldsymbol{w_m}$ denotes row m of the matrix $\boldsymbol{W}$ and where $\boldsymbol{Z}$ denotes the $T \times 1$ column vector with coordinates Z_t. (From the computational standpoint, we're calculating the integral required to convert a continuous- to a discrete-time SIR model [40].) In vector notation, our least squares criterion becomes:

$$V = (\boldsymbol{y} - \boldsymbol{WZ})'(\boldsymbol{y} - \boldsymbol{WZ}) \tag{25}$$

Let $\boldsymbol{DZ}$ denote the $T \times 3$ matrix whose row t is the gradient $\boldsymbol{DZ_t}$, and let $\boldsymbol{W(DZ)}$ denote the $M \times 3$ matrix derived by multiplying the $M \times T$ matrix $\boldsymbol{W}$ by the $T \times 3$ matrix $\boldsymbol{DZ}$. The gradient of V is then given by:

$$\boldsymbol{DV} = -2\ (\boldsymbol{y} - \boldsymbol{WZ})'\boldsymbol{W(DZ)} \tag{26}$$

Let $\boldsymbol{w_m}$ denote row m of the matrix $\boldsymbol{W}$, so that $\boldsymbol{w_m Z} = \sum_{t=1}^{T} w_{mt} Z_t$ represents the predicted value. The corresponding Hessian matrix becomes:

$$\boldsymbol{D^2 V} = 2\ \big(\boldsymbol{W(DZ)}\big)'\big(\boldsymbol{W(DZ)}\big) - \sum_{m=1}^{M}(y_m - \boldsymbol{w_m Z}) \sum_{t=1}^{T} w_{mt} \boldsymbol{D^2 Z_t} \tag{27}$$

Accordingly, we can run our SIR model to generate the state variables Z_t and then calculate the derivatives $\boldsymbol{DZ_t}$ and $\boldsymbol{D^2 Z_t}$ in the disaggregated time scale and then apply the aggregation matrix $\boldsymbol{W}$ to weight the results.

Data

Omicron Wave, December 2021 – March 2022, New York City

We studied the reported daily incidence of COVID-19 during the SARS-CoV-2 Omicron/BA.1 wave of December 2021 – March 2022 in New York City, NY, United States, a city of population 8.49 million. Our data consisted of daily counts of cases reported by the New York City department of health [80], where the date of report was intended to be the date when a positive test was performed or when the diagnosis of COVID-19 was otherwise made.

Our data showed systematic variation in case counts by day of the week, with many fewer cases diagnosed over the weekends. To account for these fluctuations, we converted the raw case counts c_t into centered 7-day moving averages, that is, $y_t = \frac{1}{7}\sum_{i=-3}^{+3} c_{t+i}$, where t indexes the date of report. Figure 1 shows the raw counts of reported cases per day c_t (connected gray datapoints) as well as the daily case counts y_t adjusted for the day of the week (red datapoints). We relied upon the adjusted daily counts y_t to estimate our SIR model parameters.

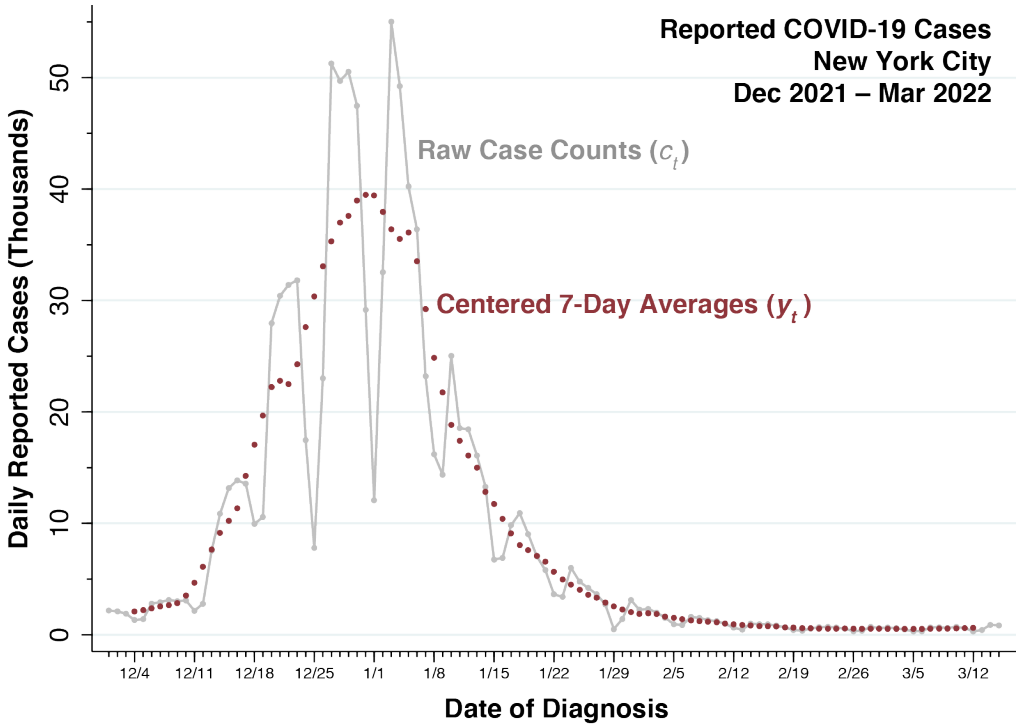

Figure 1. Daily Reported Cases of COVID-19, 12/1/2021 – 3/15/2022, in New York City. The connected gray datapoints show the raw case counts (c_t). The red datapoints, covering 12/4/2021 – 3/12/2022, show the centered 7-day moving averages (y_t).

Plague Outbreak, December 1905 – July 1906, Isle of Bombay

Figure 2 displays the average daily number of reported deaths from plague in the Isle of Bombay during each week over a 31-week period, beginning with the week starting Sunday, December 17, 1905, and continuing through the week starting Sunday, July 15, 1906. The average daily deaths were computed from the weekly totals plotted by KM in their classic 1927 paper [5]. The height of each bar is the average daily number of deaths during that week. The area of each bar equals the reported number of deaths during that week.

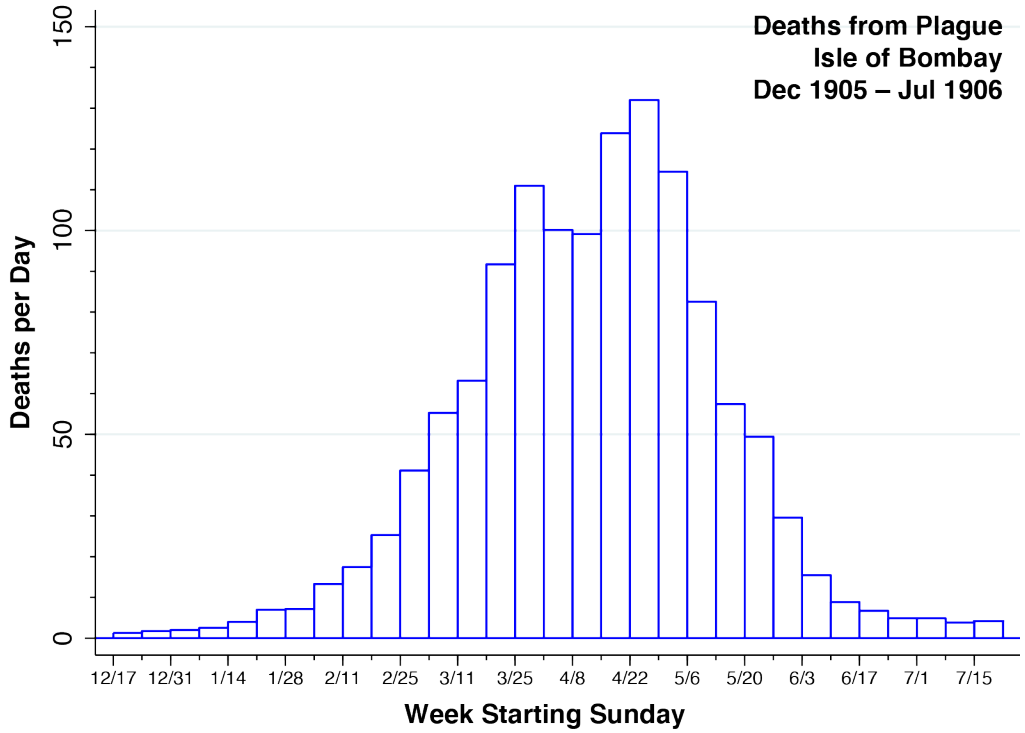

Figure 2. Average Number of Daily Deaths from Plague, Isle of Bombay, Week Starting 12/17/1905 Through the Week Starting 7/15/1906. Each bar corresponds to one week. The height of each bar is the average daily number of deaths during that week. The area of each bar equals the number of deaths during that week. Data adapted from Kermack and McKendrick [5].

In our analysis of the plague death data, we interpreted the weekly death counts y_m, indexed by $m = 1, \ldots, 31$, as observations on the aggregated values $\sum_{t=1}^{T} w_{mt} Z_t$, where Z_t are the underlying output variables with daily index $t = 1, \ldots, 217$, defined in equation (21), and where w_{mt} are elements of the aggregation matrix $\boldsymbol{W}$ described above. We estimated the parameters of our SIR model by minimizing the least squares criterion V defined in equation (25).

Data Processing and Computation

The raw COVID-19 case counts (c_t) were downloaded from the New York City data portal [80] in comma-separated-value (CSV) format and converted via Stata Statistical Software Release 17 [81] into internal Stata (DTA) format. The 7-day moving averages (y_t), displayed in Figure 1, were computed via Stata programming language [82]. The data on weekly deaths from plague in the Isle of Bombay were taken from the graph on page 714 of the original KM paper [5]. All code for parameter estimation described in the Statistical Methods section above was written in Mata [83], a matrix programming language embedded within Stata. Calculations were carried out on a MacBook Pro with a 2.3 GHz 8-Core Intel Core i9 processor.

Results

Omicron Wave, December 2021 – March 2022, New York City

Figure 3 shows the predicted values of the output variable X_t as a connected curve superimposed on the observed datapoints y_t for the Omicron wave in New York City. The computation time to achieve convergence of our Newton-Raphson algorithm was 0.93 sec.

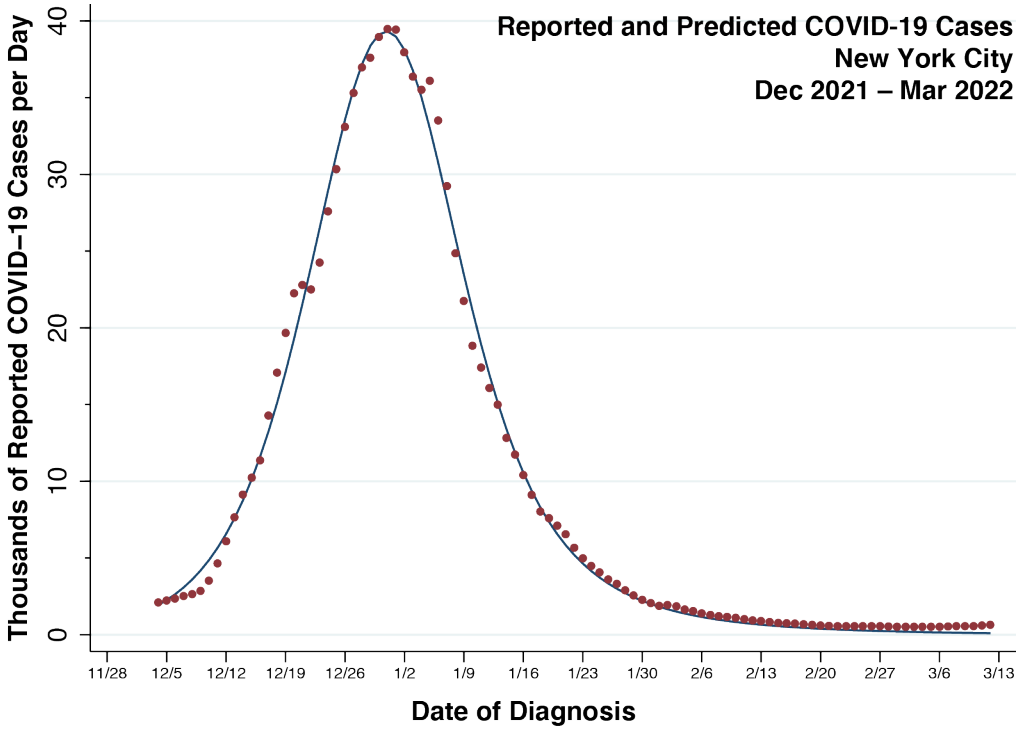

Figure 3. Daily Reported and Predicted Cases of COVID-19, 12/4/2021 – 3/12/2022, New York City. The red points show the data y_t derived from Figure 1. The curve connects the predicted values of the output variable X_t.

Table 1 summarizes the resulting parameter estimates. At $t = 0$, corresponding to December 3, 2021, the infected proportion i_0 was an estimated 0.8 percent. The estimated population-size N was about 12 percent of the city's total population. The estimated basic reproduction number $\mathcal{R}_0$ was on the order of 4. The estimated mean duration of infectivity $1/(1-\alpha)$ was in the range of 2 to 3 weeks. The 95% confidence intervals were estimated conditional upon N, as described above.

Table 1. Parameter Estimates of SIR Model of COVID-19 Incidence, New York City Omicron Wave, December 2021 – March 2022 [a,b]

β	α	$i_0 \times 10^{-3}$	$N \times 10^6$	$\mathcal{R}_0$	$1/(1-\alpha)$
0.233	0.941	8.23	1.013	3.93	16.8
(0.216, 0.250)	(0.928, 0.954)	(6.96, 9.51)		(3.35, 4.50)	(13.2, 20.5)

a. Except for the population size parameter N, symmetric 95% confidence intervals conditional upon N are displayed below each estimate.
b. In addition to the four model parameters, the final two columns show the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the estimated mean duration of infectivity $1/(1-\alpha)$.

Figure 4A plots the least squares criterion V (blue curve, left axis) and the first partial derivative $\partial V/\partial\beta$ (red curve, right axis) as functions of the parameter β. The remaining parameters have been held constant at their estimated values. The criterion V reaches a minimum at the optimum $\beta = 0.233$, where $\partial V/\partial\beta = 0$. The function V is convex in the interval from $\beta = 0.169$, where $\partial V/\partial\beta$ reaches a minimum, to $\beta = 0.352$, where $\partial V/\partial\beta$ reaches a maximum.

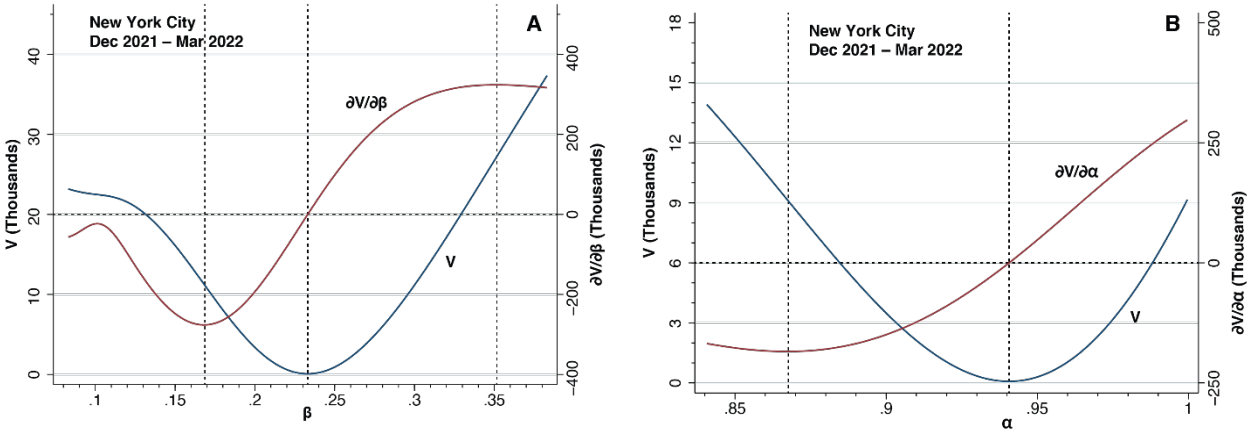

Figure 4. Panel A. Least Squares Criterion V (Left Axis) and First Partial Derivative $\partial V/\partial\beta$ (Right Axis) as Functions of the Parameter β. Panel B. Least Squares Criterion V (Left Axis) and First Partial Derivative $\partial V/\partial\alpha$ (Right Axis) as Functions of the Parameter α.

Figure 4B above displays the analogous plot of V and $\partial V/\partial\alpha$ as functions of the parameter α, where the remaining parameters are similarly held constant at their optimum values. The criterion V reaches a minimum at the optimum $\alpha = 0.941$, at which point $\partial V/\partial\alpha = 0$. The function V is convex in the interval from $\alpha = 0.868$, where $\partial V/\partial\alpha$ reaches a minimum, to $\alpha = 1$, the boundary of admissible values of α, where $\partial V/\partial\alpha$ remains positive.

Figure 5 plots the projection of V onto the (α, β) plane. Again, the remaining parameters (i_0, N) were held at the optimum values given in Table 1 above. The darkest area represents a ravine where V attains its lowest values. The yellow point in the center is the optimum $(\alpha, \beta) = (0.941, 0.233)$ where V is minimized.

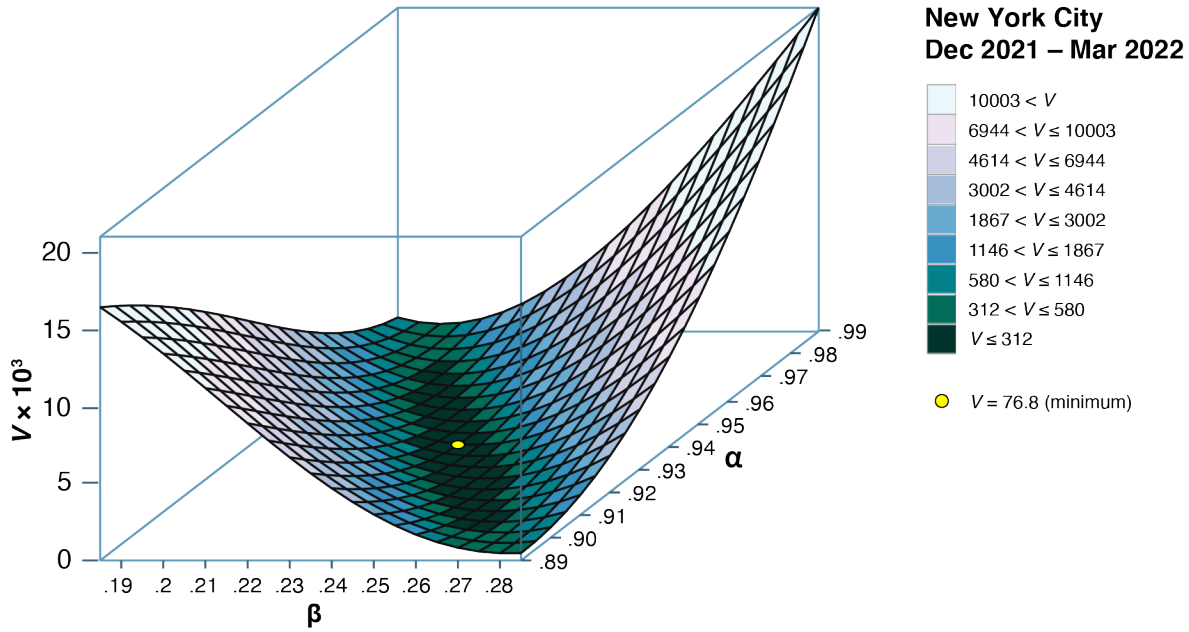

Figure 5. Least Squares Criterion V Projected onto the (α, β) Plane. The yellow point identifies the minimum where $(\alpha, \beta) = (0.941, 0.233)$. The parameters i_0 and N were held constant at their optimum values given in Table 1. The plot was produced in part from the Stata program *surface* [84].

Plague Outbreak, December 1905 – July 1906, Isle of Bombay

Figure 6 reports our analysis of data on deaths from plague during December 1905 – July 1906 in the Isle of Bombay, originally reported and studied by KM [5].

As in Figure 2 above, we have displayed the raw data on deaths as a bar graph, where each vertical bar covers one week, indexed $m = 1, \ldots, 31$, where the height of each bar is the reported average daily deaths that week, and where the area of each bar is the reported number of

weekly deaths (y_m). The superimposed green curve represents the fit to the data, based upon the KM's approximation $\frac{800}{7}\,\mathrm{sech}^2(0.2t - 3.4)$, where sech represents the hyperbolic secant, and where we have divided their published formula by 7 to convert their estimates into daily units. The green curve thus corresponds to the authors' estimates of the underlying output variable Z_t.

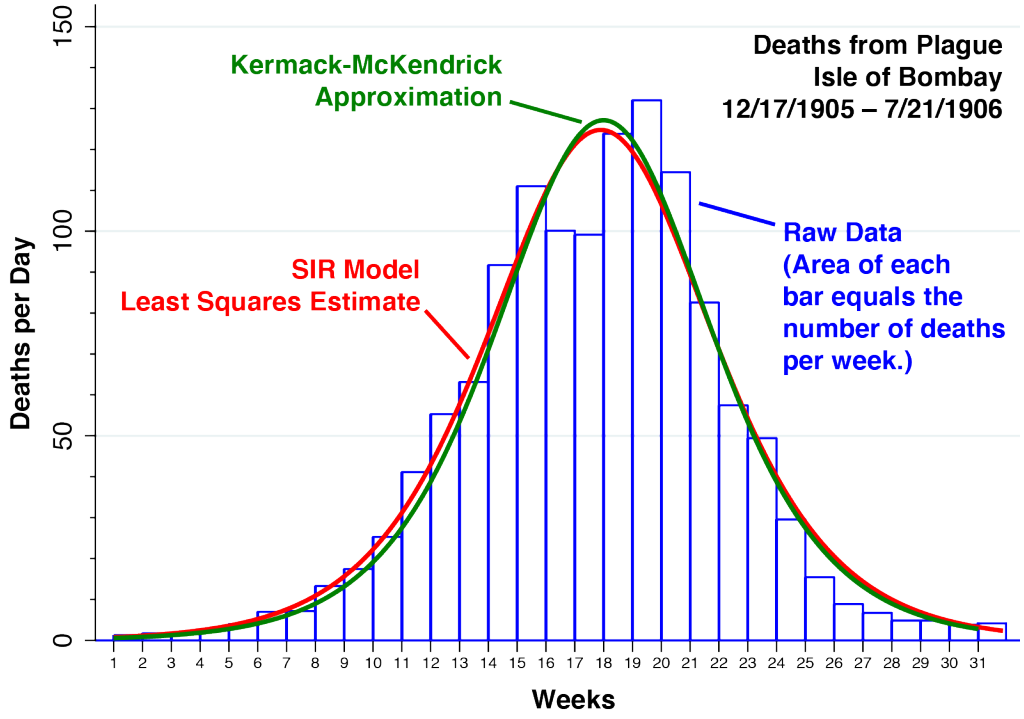

Figure 6. Reported and Predicted Deaths from Plague, Isle of Bombay, 12/17/1905 – 7/21/1906. The time axis is marked off in weeks, with week $m = 1$ starting on 12/17/1905 and week $m = 31$ starting on 7/15/1906. As noted in Figure 2, the height of each bar is the average daily number of deaths reported each week, while the area of each bar equals the number of deaths during that week, adapted from Kermack and McKendrick [5]. The green curve represents the authors' original fit to the data, while the red curve represents our estimates of the output variable Z_t.

The red curve in Figure 6 connects our predictions of the daily values Z_t for $t = 1, \ldots, 217$. The underlying parameter estimates were: $\beta = 0.610$, with 95% confidence interval (0.537, 0.683); $\alpha = 0.446$ (0.376, 0.515); $i_0 = 2.80 \times 10^{-5}$ $(1.26 \times 10^{-5}, 4.34 \times 10^{-5})$; and $N = 5.103 \times 10^4$. The estimated basic reproduction number was $\mathcal{R}_0 = 1.101$ (1.094, 1.107), while the estimated mean duration in the infected state was $1/(1-\alpha) = 1.804$ (1.579, 2.031).

Additional Results

In addition to our main results above, we performed various ancillary analyses, which are reported in appendices. Appendix B provides an approach to computing the unconditional

confidence intervals around the parameter estimates for the New York City Omicron wave. In Appendix C, we display the two- and three-dimensional paths of successive parameter estimates $\boldsymbol{\theta}^{(k)} = \left(\beta^{(k)}, \alpha^{(k)}, i_0^{(k)}\right)$ derived from our application of the Newton-Raphson algorithm to the New York City data. In Appendix D, we report the results of a robustness test, where we re-estimate our SIR model for New York City after multiplying the original case counts y_t by mean-preserving, lognormally distributed noise.

In Appendix E, we account for the observation that some of the SARS-CoV-2 cases reported during the rising phase of the New York City epidemic wave of Figure 1 were in fact infections by the Delta variant, while some of the cases reported during the declining phase were in fact Omicron BA.2 infections. To that end, we reran our SIR model after multiplying the original case counts y_t by the estimated proportion of Omicron BA.1 cases at each date t. We also reran our SIR model on the original case counts y_t, but restricted the observation interval to those dates t during which the BA.1 variant was estimated to comprise at least 80% of all cases.

In Appendices F and G, we report estimates of our SIR model for two other jurisdictions during the COVID-19 epidemic. Appendix F shows our estimates for the SARS-CoV-2 Omicron wave in Los Angeles County, CA, during the same time period covered by our study of New York City above. Appendix G shows our estimates based upon an outbreak of SARS-CoV-2 at the University of Wisconsin-Madison during September 2020, originally reported in a study of the potential super-spreader influence of a nearby cluster of local bars [79].

Finally, Appendix H provides a diagnostic plot of the least squares criterion V as a function of the parameter β for the Isle of Bombay plague data, comparable to Figure 4A for New York City above.

Discussion

Is SIR the Correct Structural Model for Omicron?

The mere fact that we have found a workable solution to the SIR inversion problem does not necessarily mean that the Susceptible-Infected-Removed model is the most appropriate description of the data under study. Our illustrative analysis of the SARS-CoV-2 Omicron/BA.1 wave in New York City during December 2021 – March 2022 cogently brings home the point.

Figure 3 shows a tight fit of the SIR model-predicted curve X_t to the case incidence datapoints y_t for New York City. Figures 4A, 4B, and 5 confirm that the estimated values of the

parameters β and α are indeed situated at the minimum of a convex region of our least squares criterion function V. Appendix C confirms that successive updates of the parameter vector $\boldsymbol{\theta}^{(\boldsymbol{k})}$ generated by the Newton-Raphson algorithm follow a path along the convex surface of V. Appendix D confirms that our estimates are robust to the inclusion of noise in the observed data.

Despite these indicators that the underlying model was not somehow ill-conditioned, the mean duration of infectivity, computed as $1/(1-\alpha)$, was estimated from the New York City data to be on the order of 17 days, with a 95% confidence interval of two to three weeks. That's way out of line with what is known from direct clinical measurement. In one cohort of 55 symptomatic Omicron-infected patients, only 13.5 percent continued to shed virus ten days after infection [85]. Yet our New York City-based parameter estimate for α would give the proportion remaining infectious after ten days at $\alpha^{10} = 54.4\%$.

To be sure, some of our alternative analyses yielded lower estimates of the mean duration of infectivity. When we attempted to exclude Delta and BA.2 infections from the original New York City case counts y_t, the estimated mean duration dropped to about 10 days (Appendix E). When we reran our SIR model on the Omicron wave in Los Angeles County, the estimated mean duration was about 9 days (Appendix F). But these estimates are still don't come close to those based on the available clinical data. When we reran our SIR model on an outbreak at the University of Wisconsin-Madison in September 2020, we obtained an estimated mean duration of about 7 days (Appendix G). But that outbreak was quite likely driven by the ancestral strain of SARS-CoV-2, and certainly not by Omicron.

Despite its remarkable fit to the data, SIR may thus fail as a *structural* model of the Omicron wave, even if it apparently succeeds as a *reduced form* model [86]. While its parameters β and α can indeed be backed out from the data and the inversion problem solved, these parameters do not necessarily warrant the structural interpretation that we assumed in our exposition of the SIR model in equations (1) through (5) above. That is not to say, however, that an adequate reduced form model is incapable of making accurate projections [87].

We should not, however, abandon the possibility that SIR may indeed be the correct structural model of the Omicron wave. To the contrary, the parameter estimates may reveal a critically important characteristic of the Omicron variant that clinical studies, which are necessarily biased toward symptomatic patients, have so far missed. In effect, the Omicron wave was so massive not because the variant had such a high per-contact infectivity (through the

parameter β), but because a much larger proportion of asymptomatic infected individuals remained persistently infectious (through the parameter α). We also need to consider that the degree of persistent infectivity is not determined solely by biological characteristics of the virus, but also by the extent to which infected individuals isolated themselves from others. That behavioral component of the parameter α may have changed during the Omicron wave.

It might be argued that there is a third plausible interpretation, namely, that SIR may indeed be the correct structural model, but that the case incidence data analyzed here were inadequate to separately identify the key parameters β and α. That is, our four-parameter version of SIR is what some physicists have described as a "sloppy" model [53, 55]. Indeed, the dark ravine in the surface map of V in Figure 5 points to a high correlation between the two parameters in the topological neighborhood of the global optimum in the (β, α) plane, an observation that has been noted previously [71]. That finding suggests that the basic reproduction number $\mathcal{R}_0 = \beta/\gamma = \beta/(1-\alpha)$ is the only identifiable or "stiff" parameter.

Still, the results in Appendix C, which follows the path of the Newton-Raphson search algorithm in parameter space, suggest that β and α are indeed separately identified from the available count data. In Figure C2, as the algorithm approaches convergence to the minimum of the criterion V, the implicit derivative $d\alpha^{(k)}/d\beta^{(k)}$ does not approach a constant value, as we would expect in the case of non-identification.

What About the Results for the Plague?

Not only do our SIR predictions fit reasonably well to the observed data on Bombay plague deaths in Figure 6, but they virtually coincide with the curve of the approximate closed-form solution drawn by KM [5]. While others have fitted SIR-type compartmental models to the 1905–1906 Bombay plague data [88], our study appears to be the first replication of the authors' century-old result.

The plague, caused by the bacterium *Yersinia pestis*, can spread through human populations by various transmission pathways. Bubonic plague, generally thought to be the cause of a series of Bombay outbreaks that began in 1896 [89-91], is transmitted to humans through rat fleas [92], but transmission between humans also appears to occur via infected human fleas or body lice [93]. Pneumonic plague, by contrast, is transmitted by direct human-to-human transmission via respiratory droplets and can occur as a complication of bubonic plague [94, 95].

Pneumonic plague is more lethal, but transmission requires closer person-to-person contact and is thus thought to have a lower basic reproduction number $\mathcal{R}_0$ [94].

So, does our SIR model similarly succeed as a reduced form model but fail as a structural model of the Bombay plague outbreak? Several structural compartmental models of plague transmission have been tested, typically involving separate states and parameters for humans and vectors [88, 89, 95-97]. In a rat-flea transmission model for bubonic plague, for example, the structural model contained a separate sub-epidemic module for rats with a distinct basic reproduction number for rat-to-rat transmission [97]. A human-flea-human structural model appeared to have the best fit to nine plague outbreaks during Europe's Second Pandemic from 1348 to 1813 [95]. Our minimalist SIR model has none of these structural features.

These general observations, however, hardly exclude the possibility that the pneumonic form may have been responsible for the 1905-1906 outbreak studied by KM. In his classic 1926 *Treatise on Pneumonic Plague* [98], Wu Lien-Teh acknowledged the 1908 Indian Plague Commission's conclusion that pneumonic constituted less than 3 percent of cases overall. (See [98] at p. 9.) But he did not hesitate to point out that some "well-characterized pneumonic outbreaks are on record." (p. 90-92.) "In fact," he added, "it was this great incidence of respiratory diseases in Bombay which attracted Childe's attention and led to his finally establishing the pneumonic form of plague as a distinct entity." (pp. 11-12). He further noted an observation bias in favor of bubonic plague due to the rapid mortality from pneumonic "…because, as a rule, such patients succumb so quickly that they seldom reach the hospital." (p. 269).

What's more, the parameter estimates derived from our solution to the inversion problem are strikingly consistent with what is known about pneumonic plague transmission. Lien-Teh's data suggested a median duration of illness of about two days from the onset of symptoms to inevitable death from heart failure ([98] at p. 249, 260). Our point estimate of the parameter $\alpha = 0.446$, implying an average infectious period of 1.8 days, is similarly consistent with more contemporary estimates of the parameters of pneumonic human-to-human transmission [95]. Lien-Teh further stressed a very low early infectivity due to the initial lack of cough (p. 296). Transmission through contact at close range, he stressed, appeared to have been required (p. 299). His observations, along with more contemporary estimates of pneumonic human-to-human transmission [94, 95], are consistent with our estimate of a basic reproduction number $\mathcal{R}_0 = 1.1$.

The plague may indeed have been initially imported into Bombay in its bubonic form through rat-flea-human transmission, and the disease may have been sustained between outbreaks either by repeated importations or by an urban reservoir of selectively immune rats [89]. Still, our data are consistent the intriguing hypothesis that responsive isolation measures by British authorities [90] may have altered the mode of transmission during the 1905–1906 outbreak from the bubonic rat-flea-human mode to the pneumonic human-human mode, a phenomenon observed in the 17th century when the Derbyshire village of Eyan fell victim to the Black Death [99]. If so, our parsimonious SIR model is in fact the appropriate structural model of the outbreak.

Why are the Estimated Population Sizes N So Small? .

Table 1 reports estimated population size parameters of $N = 1.013$ million for New York City, while Appendix Table F1 gives $N = 1.046$ million for Los Angeles County. As noted, these estimates of N were only 11.9 percent of the total census population of New York City and 14.5 percent of the corresponding census population of Los Angeles County. Where, then, did all the other millions of inhabitants go?

As discussed above, we cannot distinguish between three possible explanations of the shortfall. First, there is concrete evidence of significant underreporting of Omicron infections, due principally to the widespread availability of home rapid antigen testing [69], particularly in New York City [100, 101]. Second, despite the evidence of immune escape on the part of the Omicron variant [102], a substantial fraction of the population may still have retained long-term cellular immunity, particular through the administration of multiple vaccine doses [103]. Third, contrary to the underlying assumption of homogeneous mixing, there is strong evidence that a substantial proportion of the population avoided retail establishments, drinking and eating places, transportation venues, worksites and other high-risk locations [104].

Our analysis of the Bombay outbreak data gave an estimated size parameter $N = 51$ thousand individuals, which comes to about 6.5 percent of estimated Bombay population of 775 thousand in 1900 [105]. The same three factors could well be responsible for the shortfall. In particular, there is significant evidence for the acquisition of at least humoral immunity against *Yersinia pestis* among plague survivors [106].

Limitations and Extensions

Parameter search algorithms such as Newton-Raphson [46] and Gauss-Newton [47] have superior performance when they rely on closed-form expressions for the gradient vector and matrix of second derivates, rather than on numerical approximation [45]. Still, as Figures 4A and 4B teach us, there will nonetheless be a restricted region of the parameter space where the least squares criterion function is convex. Some of these regions may contain local rather than global optima. To avoid such problems of implementation, we plotted the least squares criterion V and its first partial derivatives as functions of the parameters. Still, finding the right initial values and keeping the parameter search within bounds remain unavoidable challenges.

As seen in Figure 1A and Appendix Figures F1 and G1, striking day-of-the-week effects on COVID-19 testing and case reporting required us to pretreat the raw case counts. Our centered, 7-day moving average appeared to be the most flexible nonparametric approach to this task. Alternative approaches that incorporate parametric models of day-of-the-week effects directly in our model of the output variable X_t have yet to be tried.

We have focused sharply on the original SIR model, rather than its numerous variations. Still, our basic approach can be extended to these more complex models. Our findings offer a caution, however, that such models as SEIR, which require an additional parameter governing the transition from an intermediate *exposed* state to the *infected* state, may very well turn out to be sloppy [71].

One exemption may be the well-studied SIRS model [59, 107], where individuals in the R (recovered) state can transition back to the S (susceptible) state as a result of waning immunity. The SIRS model still has three states, and the definitions of the output variables X_t and Z_t in equations (3) and (21), respectively, remain unchanged, but the parameter vector $\boldsymbol{\theta}$ has an additional component governing the rate of transition from R back to S. Since the SIRS model is known to admit oscillations with a stable endemic equilibrium [59, 108], we would be in a position to run our parameter recovery algorithm on multiple waves of data y_t. That is a task for future research.

Appendix A. Poisson Likelihood Function

While we relied upon the nonlinear least squares criterion V, as defined in equation (5) in the main text, to develop our estimation strategy, we could have employed other optimization

criteria. To illustrate, we consider the case where the observed observations $\boldsymbol{y}$ consist of count data on the number of reported cases, and assume that each observed data point y_t is independently Poisson distributed [71] with mean X_t. The joint likelihood of combined observations will be:

$$\mathcal{L} = \prod_{t=1}^{T} \frac{1}{y_t!} e^{-X_t} {X_t}^{y_t} \tag{A1}$$

As before, we assume that each output variable X_t is a function of the vector of underlying parameters $\boldsymbol{\theta}$. The log likelihood will then be:

$$\ell(\boldsymbol{\theta}) = -\sum_{t=1}^{T} X_t(\boldsymbol{\theta}) + \sum_{t=1}^{T} y_t \log X_t(\boldsymbol{\theta}) - K \tag{A2}$$

In equation (A2), the additional term $K = -\log(y_t!)$ does not depend on the underlying parameters $\boldsymbol{\theta}$. The gradient of the log likelihood function will then be:

$$\boldsymbol{D\ell} = -\sum_{t=1}^{T} \boldsymbol{DX_t} + \sum_{t=1}^{T} \frac{y_t}{X_t} \boldsymbol{DX_t} = \sum_{t=1}^{T} \frac{1}{X_t} (y_t - X_t) \boldsymbol{DX_t} \tag{A3}$$

The gradient expressed in (A3) differs from the gradient in equation (6) only in terms of the weighting factors $\frac{1}{X_t}$ in each summation term. The Hessian matrix of second-order derivatives is:

$$\boldsymbol{D^2\ell} = \sum_{t=1}^{T} \frac{1}{X_t} (y_t - X_t) \boldsymbol{D^2 X_t} - \sum_{t=1}^{T} \left(\frac{y_t}{X_t^2}\right) (\boldsymbol{DX_t}) \cdot (\boldsymbol{DX_t}) \tag{A4}$$

While the Hessian matrix in equation (10) above was positive definite reflecting an objective function V to be minimized, here the Hessian matrix is negative definite reflecting an objective function ℓ to be maximized. The remaining computations of the gradients $\boldsymbol{DX_t}$ will be computed iteratively exactly as described in the main text.

Given these expressions for the gradient $\boldsymbol{D\ell}$ and Hessian $\boldsymbol{D^2\ell}$ of the likelihood function, we can then use the Newton-Raphson algorithm [46] to find maximum likelihood estimates.

Appendix B. Computation of Unconditional Confidence Intervals: New York City Omicron

In the main text, we estimated the variance-covariance matrix of the parameters $\boldsymbol{\theta} = (\beta, \alpha, i_0)$ *conditional* on N as $\boldsymbol{C} = s^2(\boldsymbol{DX}'\boldsymbol{DX})^{-1}$, where $s^2 = V/T$, where $\boldsymbol{DX}$ was the $T \times 3$ matrix whose rows were the gradient vectors $\boldsymbol{DX_t}$ defined in (7), and where $\boldsymbol{DX_t}$ and V were evaluated at the optimum. In this Appendix, we offer one possible approach to computing the

unconditional variance-covariance matrix of the combined parameters $(\boldsymbol{\theta}, N)$ and display the comparative results for the data on the 2021-2022 Omicron wave in New York City.

We again rely upon [48] to compute the *unconditional* variance-covariance matrix $\boldsymbol{C} = s^2\left(\boldsymbol{D\widetilde{X}}'\boldsymbol{D\widetilde{X}}\right)^{-1}$, where $s^2 = V/T$ is estimated as above, but where $\boldsymbol{D\widetilde{X}}$ is the $T \times 4$ matrix whose rows are the augmented gradient vectors $\boldsymbol{D\widetilde{X}_t}=\left(\boldsymbol{DX_t}, \frac{\partial X_t}{\partial N}\right)$. In contrast to equation (7) in the main text, we compute:

$$\boldsymbol{D\widetilde{X}_t} = S_{t-1}I_{t-1}\left(\frac{1}{N}\boldsymbol{D\beta} - \frac{\beta}{N^2}\boldsymbol{DN}\right) + \left(\frac{\beta}{N}\right)I_{t-1}\boldsymbol{DS_{t-1}} + \left(\frac{\beta}{N}\right)S_{t-1}\boldsymbol{DI_{t-1}} \quad \text{(B1)}$$

Here, we define $\boldsymbol{D\beta} = (1,0,0,0)'$ and $\boldsymbol{DN} = (0,0,0,1)'$ as unit vectors in four dimensions. Equation (B1) is used only to evaluate the augmented matrix $\boldsymbol{D\widetilde{X}}$ and was not used to estimate the optimum values of the parameters. The quantities $\boldsymbol{D\widetilde{X}_t}$ and V continue to be evaluated at the optimum values of the parameters, which remain unchanged from the main text.

Table B1 compares the conditional and unconditional confidence intervals.

Table B1. Conditional and Unconditional 95% Confidence Intervals of the Parameter Estimates of the SIR Model of COVID-19 Incidence: New York City Omicron Wave, December 2021 – March 2022 [a]

Parameter	Estimate	Conditional 95% CI	Unconditional 95% CI
$\boldsymbol{\beta}$	0.233	(0.216, 0.250)	(0.200, 0.266)
$\boldsymbol{\alpha}$	0.941	(0.928, 0.954)	(0.912, 0.969)
$\boldsymbol{i_0 \times 10^{-3}}$	8.23	(6.96, 9.51)	(6.19, 10.27)
$\boldsymbol{N \times 10^6}$	1.013		(0.980, 1.046)
$\boldsymbol{\mathcal{R}_0}$	3.93	(3.35, 4.50)	(2.59, 5.26)
$\boldsymbol{1/(1-\alpha)}$	16.8	(13.2, 20.5)	(8.75, 24.9)

a. In addition to the four model parameters, the final two rows show the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the estimated mean duration of infectivity $1/(1-\alpha)$.

Appendix C. Search Path of the Newton-Raphson Algorithm: New York City

In Figure 5 of the main text, we mapped the least squares criterion V as a function of the parameters (β, α). The optimum point, we found, was situated along a ravine where β and α appeared to be highly correlated. Here, we study the path of parameters $\boldsymbol{\theta^{(k)}}$ during successive iterations of the Newton-Raphson algorithm described in equation (15) of the main text.

Figure C1 shows the projection of the path of $\boldsymbol{\theta}^{(k)}$ onto the (β, α) plane. Rather than apply the EM algorithm to estimate the population-size parameter N, as described in equation (18) of the main text, we conditioned on a fixed $N = 1.013 \times 10^6$. We chose starting values $\boldsymbol{\theta}^{(0)} = \left(\beta^{(0)}, \alpha^{(0)}, i_0^{(0)}\right) = (0.3, 0.8, 0.009)$, corresponding to point A. The step size was $q = 0.1$.

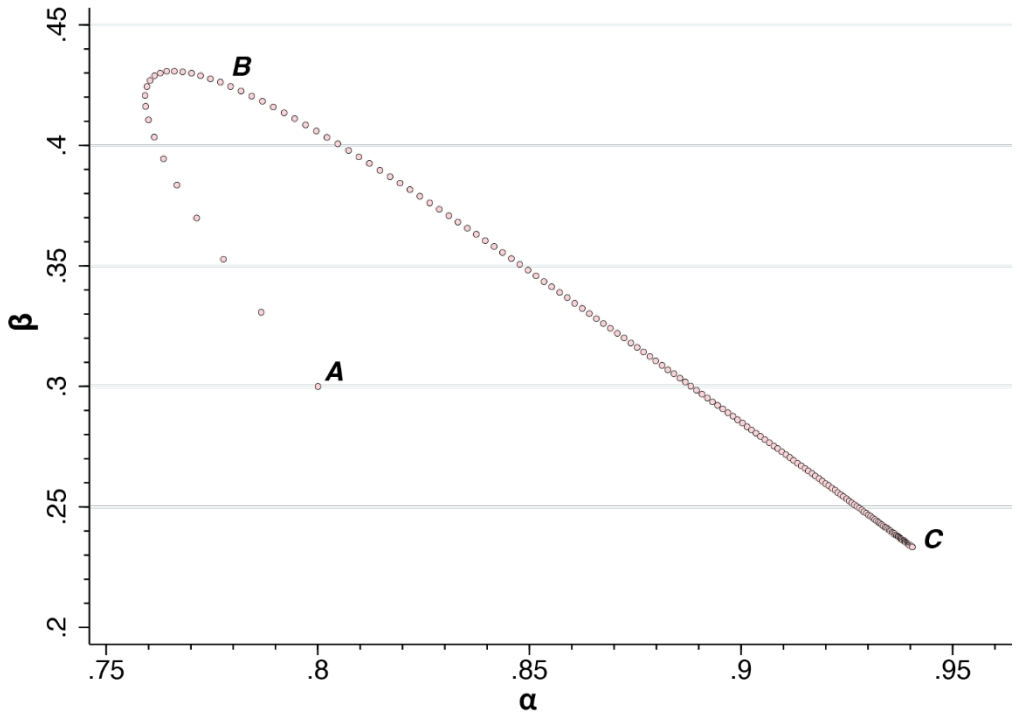

Figure C1. Path of the Parameter Vector $\boldsymbol{\theta}^{(k)}$ Projected onto the $(\boldsymbol{\beta}, \boldsymbol{\alpha},)$ Plane Through Successive Iterations of the Newton-Raphson Algorithm: New York City Omicron Wave. The results are conditional upon $N = 1.013 \times 10^6$. The algorithm started at point $A = (0.3, 0.8)$. At point $B = (0.424, 0.779)$, the path reverses and follows along the ravine described in Figure 5 of the main text, ultimately reaching convergence at point $C = (0.233, 0.940)$.

From the starting point A, the algorithm initially proceeded in the direction of increasing β and decreasing α. By point B, the algorithm had reversed course and began to follow along the ravine described in Figure 5 of the main text, ultimately reaching convergence at point $C =$ (0.233, 0.940). At each iteration, the matrix $\boldsymbol{D^2V}$ was found to be positive definite.

Figure C2 below plots the ratio of the partial derivative $\partial V/\partial\beta$ to the partial derivative $\partial V/\partial\alpha$ as a function of the iteration number k during the evolution of the algorithm. This ratio is the negative of the implicit derivative $d\alpha/d\beta$ along the search path. The ratio is slightly increasing in k along the segment of the search path where $87 \leq k \leq 122$. Otherwise, the ratio continues to decline even after the path enters the ravine identified in Figure 5 of the main text.

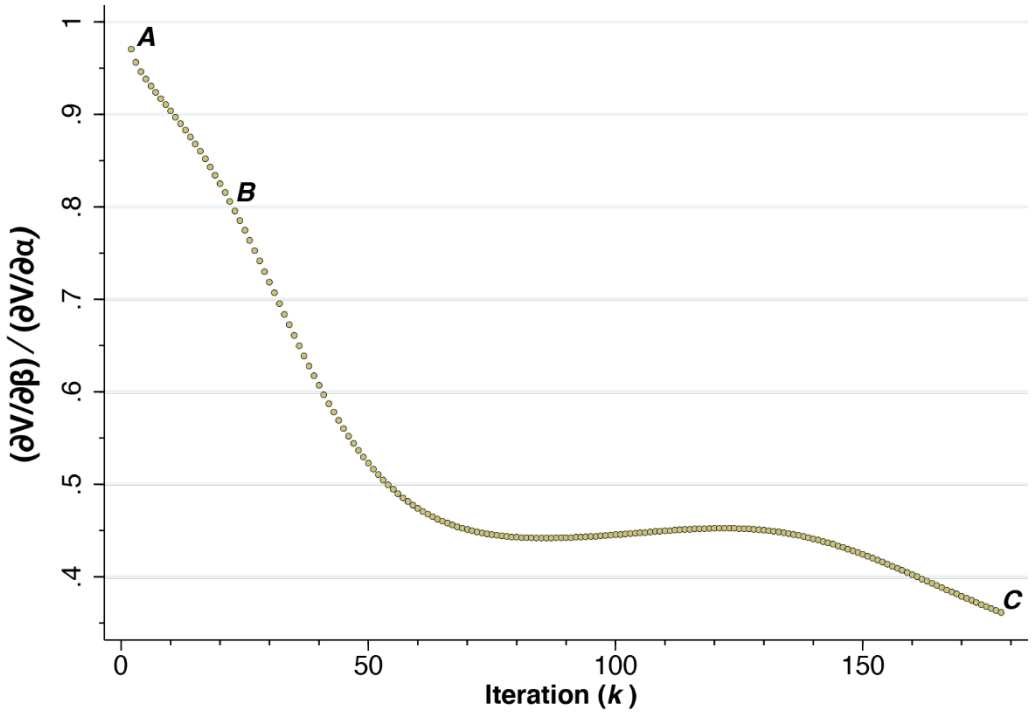

Figure C2. Path of the Ratio of $\partial V/\partial \beta$ to $\partial V/\partial \alpha$ Through Successive Iterations of the Newton-Raphson Algorithm: New York City Omicron Wave. As in Figure C1, the algorithm started at point *A* and converged at point *C*. At point *B*, the search path reversed and followed along the ravine described in Figure 5 of the main text.

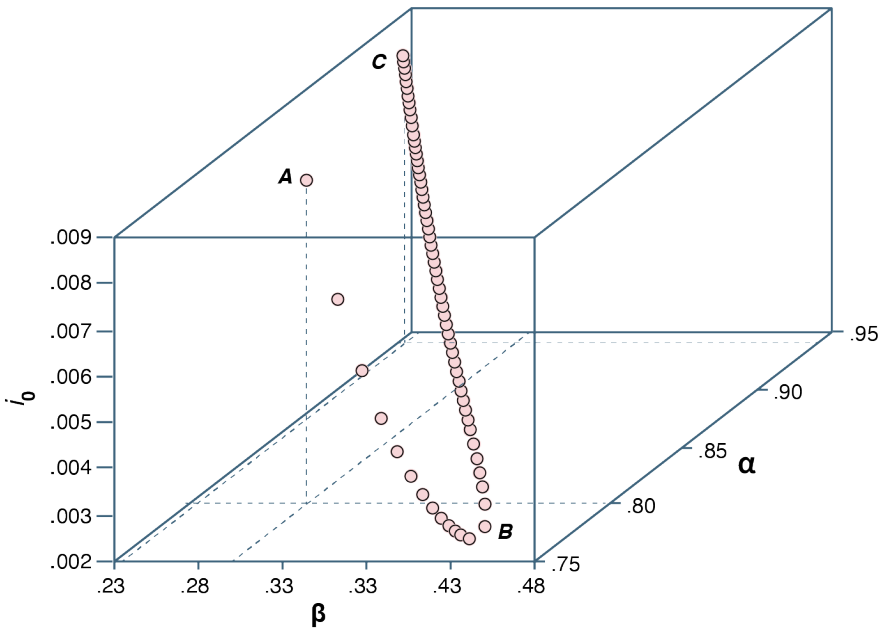

Figure C3. Path of the Parameter Vector $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{i}_0^{(k)})$ Through Successive Iterations of the Newton-Raphson Algorithm, New York City Omicron Wave. We show only those points separated in the vertical i_0 dimension by 0.0001. The algorithm started at A = (0.3, 0.8, 0.009). At B = (0.424, 0.779, 0.002), the path of $\boldsymbol{\theta}^{(k)}$ reverses direction until it converges at C = (0.233, 0.940, 0.821).

Figure C3 above shows the path of the parameters $\boldsymbol{\theta}^{(k)} = \left(\beta^{(k)}, \alpha^{(k)}, i_0^{(k)}\right)$ in three dimensions, along with the corresponding points *A*, *B*, and *C*. The search path from *B* to *C* reflects not only the movement along the ravine of Figure 5, but also the progressive increase in the estimated parameter i_0.

Appendix D. Robustness Test: New York City Omicron

Here, we report a robustness test of the parameter estimates based on the COVID-19 incidence data from New York City. To that end, we re-estimated the SIR parameters after multiplying the original incidence data by random, lognormally distributed noise.

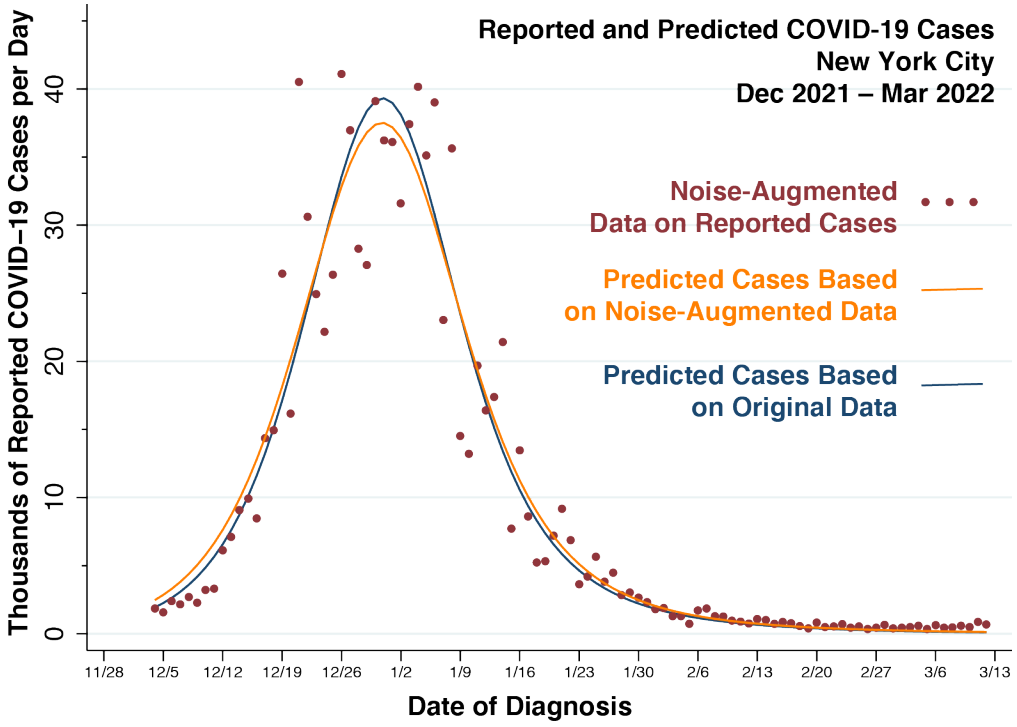

Figure D1. Noise-Augmented Data on Reported Daily COVID-19 Cases (Red Datapoints), Predicted Cases Based on Noise-Augmented Data (Orange Curve), and Predicted Cases Based on Original Data (Blue Curve), New York City, 12/4/2021 – 3/12/2022. The originally reported case counts were multiplied by independent lognormal noise with mean equal to 1 and standard deviation equal to $1/4$.

Figure D1 above plots the noise-augmented COVID-19 incidence data (red datapoints) and the resulting least-squares predicted incidence (orange curve), along with the predicted incidence based on the original data (blue curve). The noisy data were generated as $\tilde{y}_t = y_t e^{\eta_t}$, where y_t were the original data, and where the components η_t were independent draws from a

Gaussian $N(\mu, \sigma^2)$ distribution. The parameters were calibrated as $\sigma^2 = \log\left(1 + \left(\frac{1}{4}\right)^2\right) \cong 0.06062$ and $\mu = -\frac{1}{2}\sigma^2 \cong -0.03031$ so that the lognormal random variable e^{η_t} had mean $e^{\mu+\frac{1}{2}\sigma^2} = 1$ and standard deviation $\sqrt{\left(e^{\sigma^2} - 1\right)e^{2\mu+\sigma^2}} = \frac{1}{4}$.

Table D1 compares the parameter estimates based on the noise-augmented data with those derived from the original data. For these calculations, we relied on the re-parametrized $\boldsymbol{\theta} = (\beta, \omega, i_0)$ with $\alpha(\omega) = 1/(1 + e^{-\omega})$, as described in the Confidence Intervals section in the main text. The estimated confidence intervals are conditional on the estimated population size parameter N. Both Figure D1 and Table D1 indicate that the underlying model parameters β, α, i_0, and N were reasonably well conserved after the original data were multiplied by sample mean-preserving noise. As anticipated, the confidence intervals surrounding the parameter estimates were considerably wider.

Table D1. Parameter Estimates of an SIR Model of COVID-19 Incidence New York City, December 2021 – March 2022, With and Without Multiplicative Lognormal Noise [a,b]

Parameter	Original Estimates [c]	Noisy Estimates [c]
$\boldsymbol{\beta}$	0.233 (0.216, 0.250)	0.209 (0.140, 0.278)
$\boldsymbol{\omega}$	2.763 (2.532, 2.994)	3.000 (1.831, 4.162)
$\boldsymbol{\alpha}$	0.941 (0.926, 0.952)	0.952 (0.862, 0.985)
$\boldsymbol{i_0 \times 10^{-3}}$	8.23 (6.96, 9.51)	11.68 (3.84, 19.51)
$\boldsymbol{N \times 10^6}$	1.013	1.024
$\boldsymbol{\mathcal{R}_0}$	3.93 (3.35, 4.50)	4.39 (0.95, 7.82)
$\boldsymbol{1/(1-\alpha)}$	16.8 (13.6, 21.0)	21.0 (7.24, 65.2)

a. Except for the population size parameter N, 95% confidence intervals conditional upon N are displayed below each estimate. Confidence intervals were based on the re-parametrized $\boldsymbol{\theta} = (\beta, \omega, i_0)$ with $\alpha(\omega) = 1/(1 + e^{-\omega})$.

b. In addition to the underlying model parameters, the final two rows show the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha) = \beta(1 + e^{\omega})$ and the estimated mean duration of infectivity $1/(1-\alpha) = 1 + e^{\omega}$.

c. The sample mean of both the original datapoints y_t and the noise-augmented datapoints $\tilde{y}_t$ was 10.05.

Appendix E. Accounting for Different Variants of SARS-CoV-2: New York City

The Omicron BA.1 variant of the SARS-CoV-2 virus was far-and-away the most important contributor to the massive surge of reported infections observed in Figure 1 of the main text. Still, the initial phase of the surge overlapped the tail end of the prior Delta wave, while the terminal phase saw the gradual emergence of the Omicron BA.2 variant. In this Appendix, we repeat our SIR analysis on the data for the Omicron BA.1 variant alone.

Figure D1 plots the estimated daily proportions of the three variants (Delta, Omicron BA.1, and Omicron BA.2) in New York City during the period of observation of our study. Estimates of the average weekly proportions, represented by the solid datapoints, were available from a compilation of genomic tests of viral samples maintained by the New York City health department [109]. We employed the Stata *pchipolate* (piecewise cubic Hermite interpolation) routine [110] to estimate the proportions for the intervening days, as represented by the connecting curves.

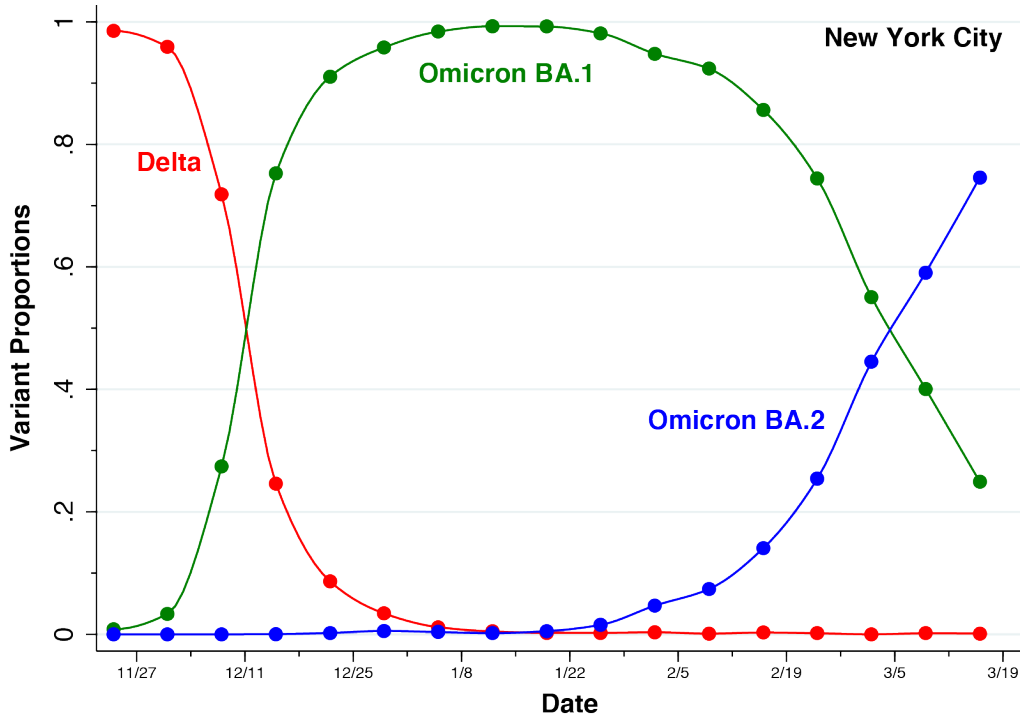

Figure E1. Estimates of the Proportions of the Delta, Omicron BA.1, and Omicron BA.2 Variants in New York City During 112/4/2021 – 3/16/2022. Weekly averages (solid datapoints) were derived from a compilation maintained by the New York City health department [109]. The Stata *pchipolate* interpolation routine [110] was then employed to estimate the intervening days (connecting curves). The Omicron BA.2 category included strains identified as BA.2, BA.2.12.1, and BA.2.75. Not shown are the proportions of all other strains, which represented no more than 0.7 percent of total samples in any one week.

Figure E2 below compares the SIR estimates derived from the Omicron BA.1 data alone (shown in orange) with the corresponding estimates reported in Figure 3 of the main text (shown in green). Elimination of the Delta cases resulted in a discernable attenuation of the initial upswing. By contrast, elimination of the Omicron BA.2 resulted in only a small absolute decrease in the tail of the wave during February and early March of 2022. By the time Omicron BA.2 began to assume a significant proportion of cases in February, as shown in Figure E1, the massive wave of cases in New York City had already dissipated.

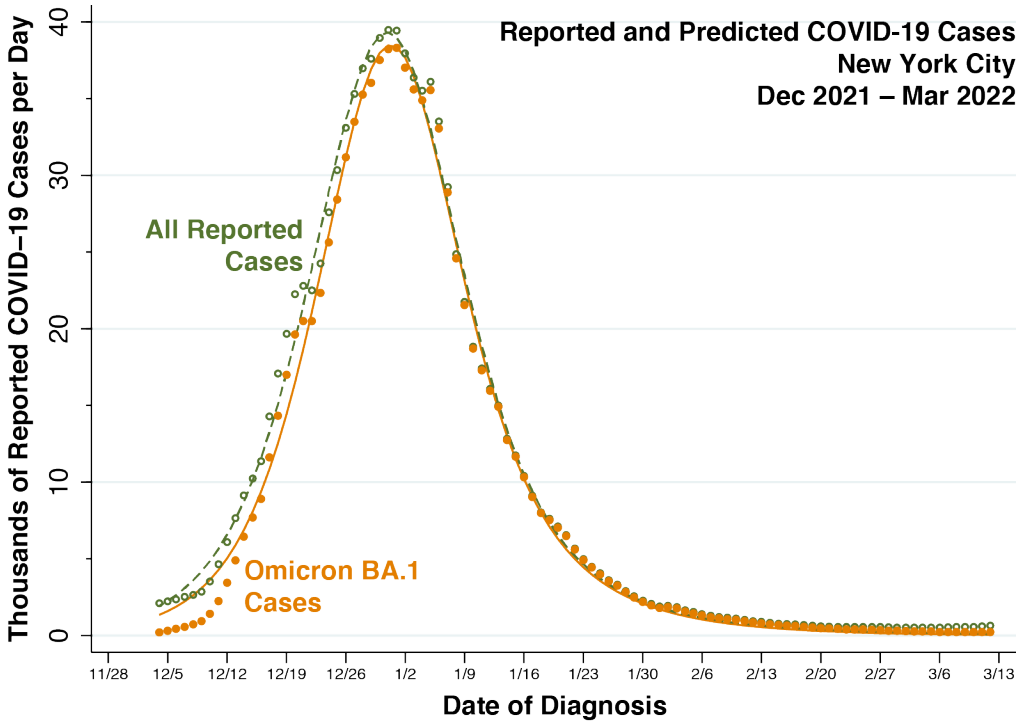

Figure E2. Daily Reported and Predicted Cases of COVID-19, 12/1/2021 – 3/15/2022, in New York City. Estimates Based Upon All Variants (Green) and Omicron BA.1 Variant Only (Orange). Reported cases are represented by data points; predicted cases by curves. Data and estimates based upon all variants reproduced from Figure 3 of main text.

Table E1 compares the original parameter estimates shown in Table 1 of the main text with the corresponding estimates derived from the Omicron BA.1-only sample. With the attenuation of the initial upswing in cases, the estimated i_0 drops by nearly half, while the basic reproduction number $\mathcal{R}_0$ declines from roughly 4 to 3. What's more, the estimated mean duration of infection falls to 10.3 days, but its 95% confidence interval still does not contain the U.S. Centers for Disease Control estimate of 5.5 days [111].

Table E1. Parameter Estimates of an SIR Model of COVID-19 Incidence New York City, December 2021 – March 2022. Data on All Variants Versus Omicron BA.1 Only [a,b]

Parameter	Original Estimates	Omicron BA.1 Only
$\boldsymbol{\beta}$	0.233 (0.216, 0.250)	0.283 (0.266, 0.299)
$\boldsymbol{\alpha}$	0.941 (0.928, 0.954)	0.903 (0.890, 0.916)
$\boldsymbol{i_0 \times 10^{-3}}$	8.23 (6.96, 9.51)	4.85 (4.15, 5.56)
$\boldsymbol{N \times 10^{6}}$	1.013	0.984
$\boldsymbol{\mathcal{R}_0}$	3.93 (3.35, 4.50)	2.92 (2.70, 3.14)
$\boldsymbol{1/(1-\alpha)}$	16.8 (13.2, 20.5)	10.32 (8.94, 11.72)

a. Except for the population size parameter N, symmetric 95% confidence intervals conditional upon N are displayed below each estimate.
b. In addition to the underlying model parameters, the final two rows show the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the estimated mean duration of infectivity $1/(1-\alpha)$.

While the Omicron BA.1-only estimates might appear more reasonable, dropping the other variants from the database is hardly innocent. To see why, consider the hypothetical case where an infection by Virus X consistently caused a false positive test for SARS-CoV-2 but conferred no protection against COVID-19. In that case, it would be appropriate to downwardly adjust the total case counts by the estimated fractions of Virus X infections.

Here, however, we cannot assume that a Delta infection confers no protection at all against a subsequent Omicron BA.1 infection. To the contrary, the Delta infections seen at the onset of the wave in Figure E1 likely reduced the proportion of individuals susceptible to Omicron BA.1 [112]. Dropping the Delta cases would downwardly bias the model estimates of S_t and thus upwardly bias the resulting estimates of the parameter β.

To avoid these potential biases, we estimated our SIR model instead on counts of all reported SARS-CoV-2 cases during the truncated 65-day window from 12/17/2021 through 2/19/2022 when the estimated proportion of Omicron BA.1 cases exceeded 80 percent. Figure E3 graphs the observed and predicted counts, while Table E2 compares the parameter estimates from the original analysis in the main text to the estimates based on the truncated data set. The results reveal strong concordance between the two sets of estimates.

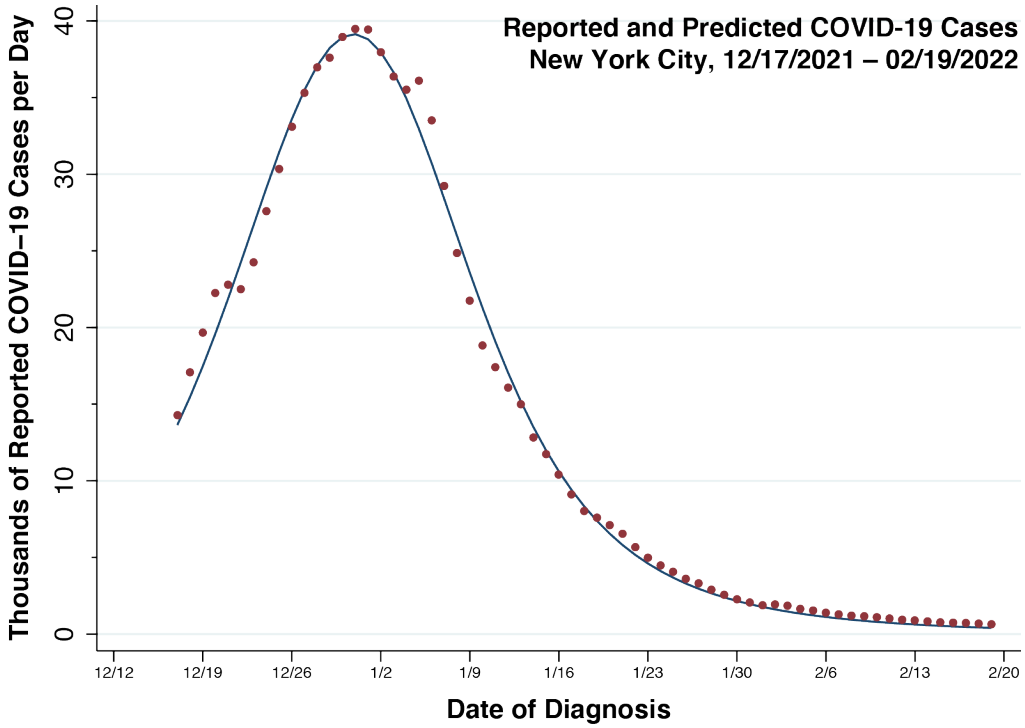

Figure E3. Daily Reported and Predicted COVID-19 Cases Based Upon Truncated Sample of New York City, 12/17/2021 – 02/19/2022. The mean squared deviation between the predicted $\check{X}_t$ based upon the 65-day truncated sample and corresponding predicted X_t based upon the full sample was $\frac{1}{65}\sum_{t=1}^{65}\left(\check{X}_t - X_t\right)^2 = 0.017$.

Table E2. Parameter Estimates of SIR Model of COVID-19 Incidence, New York City, Full Sample (12/04/2021–03/12/22) and Truncated Sample (12/17/2021–02/19/2022) [a,b]

Parameter	Original Estimates	Truncated Sample [c]
$\boldsymbol{\beta}$	0.233 (0.216, 0.250)	0.216 (0.189, 0.242)
$\boldsymbol{\alpha}$	0.941 (0.928, 0.954)	0.951 (0.929, 0.972)
$\boldsymbol{i_0 \times 10^{-3}}$	8.23 (6.96, 9.51)	68.4 (55.7, 81.8)
$\boldsymbol{N \times 10^6}$	1.013	0.993
$\boldsymbol{\mathcal{R}_0}$	3.93 (3.35, 4.50)	2.92 (2.70, 3.14)
$\boldsymbol{1/(1-\alpha)}$	16.8 (13.2, 20.5)	20.3 (11.4, 29.2)

a. Except for the population size parameter N, symmetric 95% confidence intervals conditional upon N are displayed below each estimate.
b. In addition to the underlying model parameters, the final two rows show the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the estimated mean duration of infectivity $1/(1-\alpha)$.
c. The parameter i_0 corresponds the estimated proportion infected as of 12/16/2021.

Appendix F. Omicron Wave, December 2021 – March 2022, Los Angeles County, U.S.A.

We studied the daily incidence of COVID-19 during the SARS-CoV-2 Omicron/BA.1 wave of December 2021 – March 2022 in Los Angeles County, CA, excluding the cities of Pasadena and Long Beach, a jurisdiction with population 9.26 million, as reported by the Los Angeles County Department of Public Health [113]. As in the case of New York City [80], the date of report was intended to be the date when a positive test was performed or when the diagnosis of COVID-19 was otherwise made.

Figure F1 displays the raw data for Los Angeles County, derived in the same manner as the data shown in Figure 1 of the main text for New York City. To account for systematic variation in case counts by day of day of the week, we converted the raw case counts c_t into centered 7-day moving averages, that is, $y_t = \frac{1}{7}\sum_{i=-3}^{+3} c_{t+i}$, where t indexes the date of report. Figure E1 shows the raw counts of reported cases per day c_t (connected gray datapoints) as well as the daily case counts y_t adjusted for the day of the week (red datapoints). As in the main text, we relied upon the adjusted daily counts y_t to estimate our SIR model parameters.

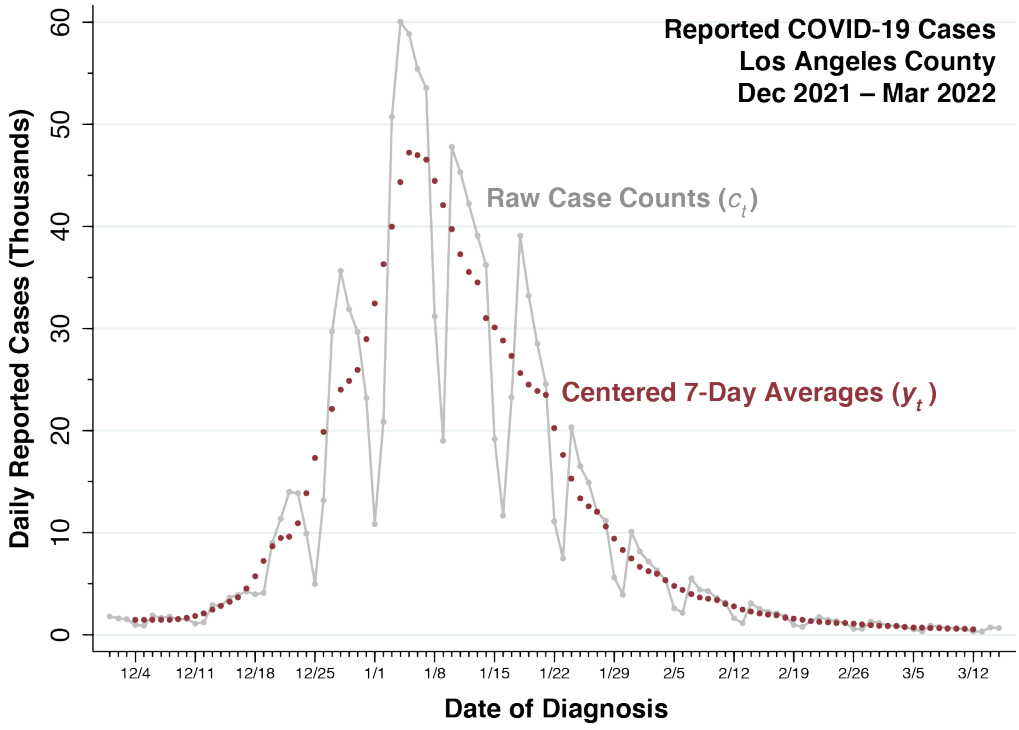

Figure F1. Daily Reported Cases of COVID-19, 12/1/2021 – 3/15/2022, in Los Angeles County. The connected gray datapoints show the raw case counts (c_t). The red datapoints, covering 12/4/2021 – 3/12/2022, show the centered 7-day moving averages (y_t).

Figure F2 shows the predicted values of the output variable X_t as a connected curve superimposed on the observed datapoints y_t for the Omicron wave in Los Angeles County. The fit of the data to the four-parameter SIR model is not as tight as for New York City, undershooting the peak reported incidence in early January 2022 by about 8 percent. The computation time to achieve convergence of the Newton-Raphson algorithm was 1.04 sec.

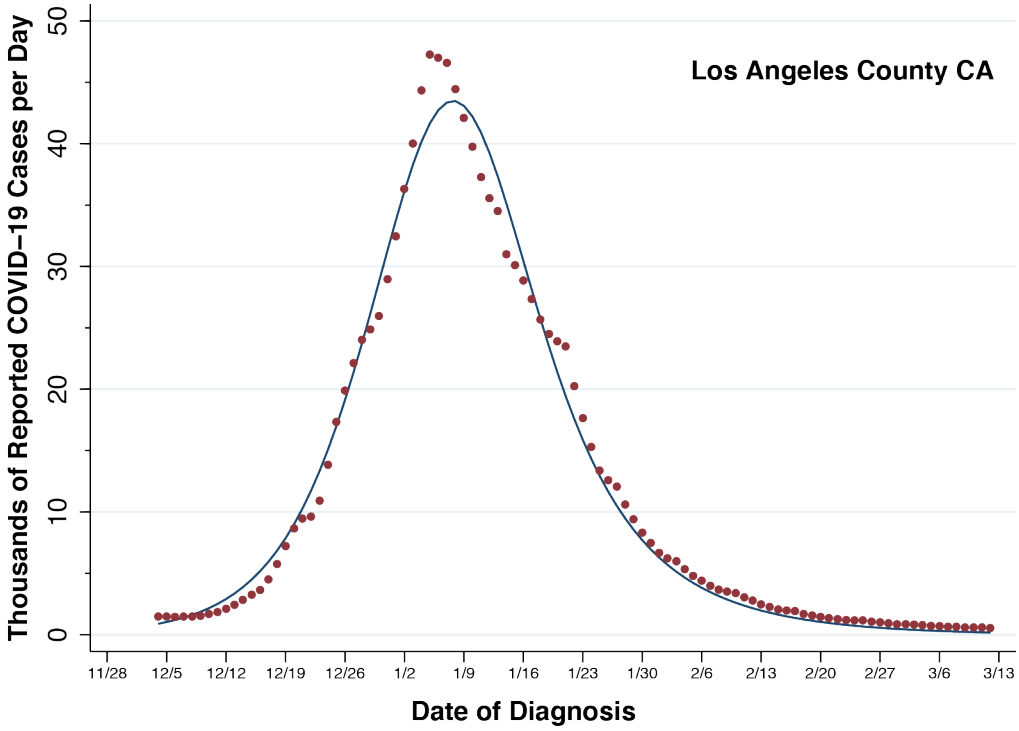

Figure F2. Daily Reported and Predicted Cases of COVID-19, 12/4/2021 – 3/12/2022, Los Angeles County. The red points show the data y_t derived from Figure F1. The curve connects the predicted values of the output variable X_t.

Table F1 summarizes the resulting parameter estimates. At $t = 0$, corresponding to December 3, 2021, the infected proportion i_0 was an estimated 0.2 percent, as compared to 0.8 percent for New York City. The estimated population-size N was about 14.5 percent of the total county population excluding the cities of Pasadena and Long Beach. The estimated basic reproduction number $\mathcal{R}_0$ was on the order of 2.5, considerably small than that estimated for New York City. The estimated mean duration of infectivity $1/(1-\alpha)$ was on the order of 9 days, likewise considerably smaller than the New York City-based estimate.

Table F1. Parameter Estimates of SIR Model of COVID-19 Incidence, Los Angeles County Omicron Wave, December 2021 – March 2022 [a,b]

β	α	$i_0 \times 10^{-3}$	$N \times 10^6$	$\mathcal{R}_0$	$1/(1-\alpha)$
0.272	0.890	2.44	1.346	2.49	9.11
(0.254, 0.290)	(0.876, 0.904)	(1.98, 2.90)		(2.32, 2.65)	(7.93, 10.3)

a. Except for the population size parameter N, symmetric 95% confidence intervals conditional upon N are displayed below each estimate.
b. In addition to the four model parameters, the final two columns show the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the estimated mean duration of infectivity $1/(1-\alpha)$.

Appendix G. University of Wisconsin-Madison COVID-19 Outbreak, September 2020.

In Figure G1, we display our SIR analysis of the data on an outbreak of SARS-CoV-2 at the University of Wisconsin-Madison during September 2020, originally reported in a study of the potential super-spreader influence of a nearby cluster of local bars [79].

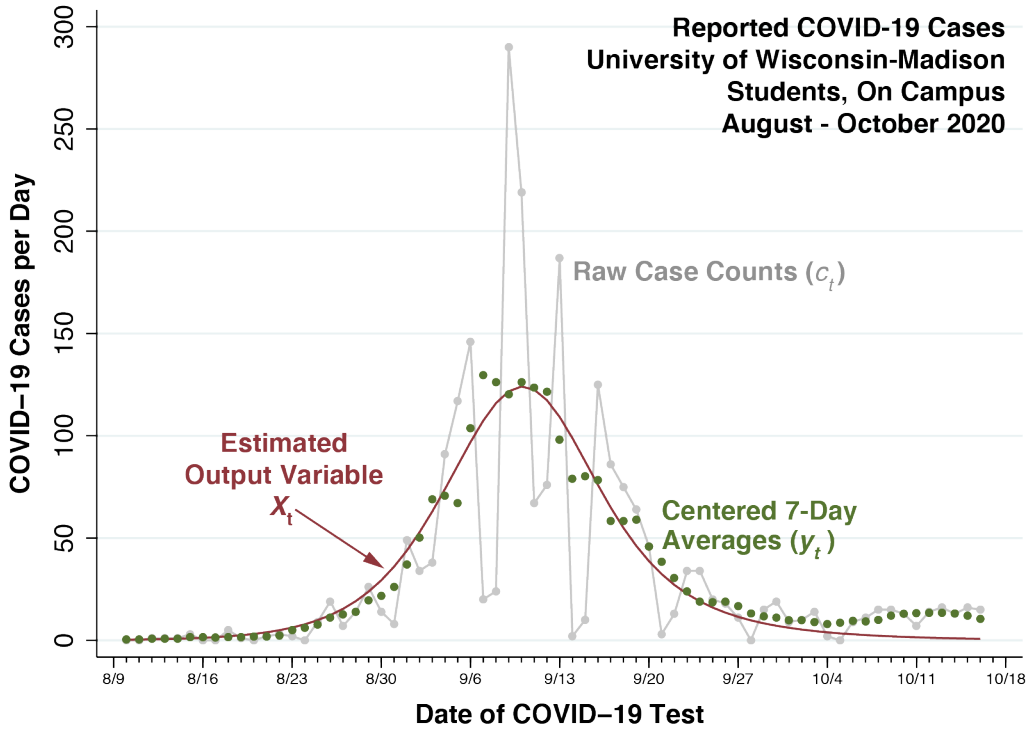

Figure G1. Raw Case Counts (c_t), Centered 7-Day Average Daily Incidence (y_t), and Estimated Output Variable (X_t) from Least Squares Fit of an SIR Model, COVID-19 Outbreak, University of Wisconsin-Madison Campus, September 2020. Source data: [79].

Table G1 displays the estimated parameters and their conditional 95% confidence intervals. The estimated number infected at the start of the wave on 8/9/2020 was $I_0 = i_0 N = 0.7$ (0.3, 1.1), which would imply that single infected individual ("Student Zero") initiated the outbreak. The estimate of N is remarkably close to the combined census of 2,392 students living in the two residence halls that had to be locked down as a result of a high prevalence of infection.

Table G1. Parameter Estimates of SIR Model of COVID-19 Outbreak, University of Wisconsin-Madison Campus, September 2020. [a,b]

β	α	$i_0 \times 10^{-3}$	N	$\mathcal{R}_0$	$1/(1-\alpha)$
0.411	0.854	0.32	2242	2.82	6.85
(0.343, 0.479)	(0.800, 0.908)	(0.15, 0.50)		(2.24, 3.40)	(4.34, 9.36)

a. Except for the population size parameter N, symmetric 95% confidence intervals conditional upon N are displayed below each estimate.
b. In addition to the four model parameters, the final two columns show the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the estimated mean duration of infectivity $1/(1-\alpha)$.

Appendix H. Diagnostic Plot for the KM Data on Deaths from Plague, Isle of Bombay

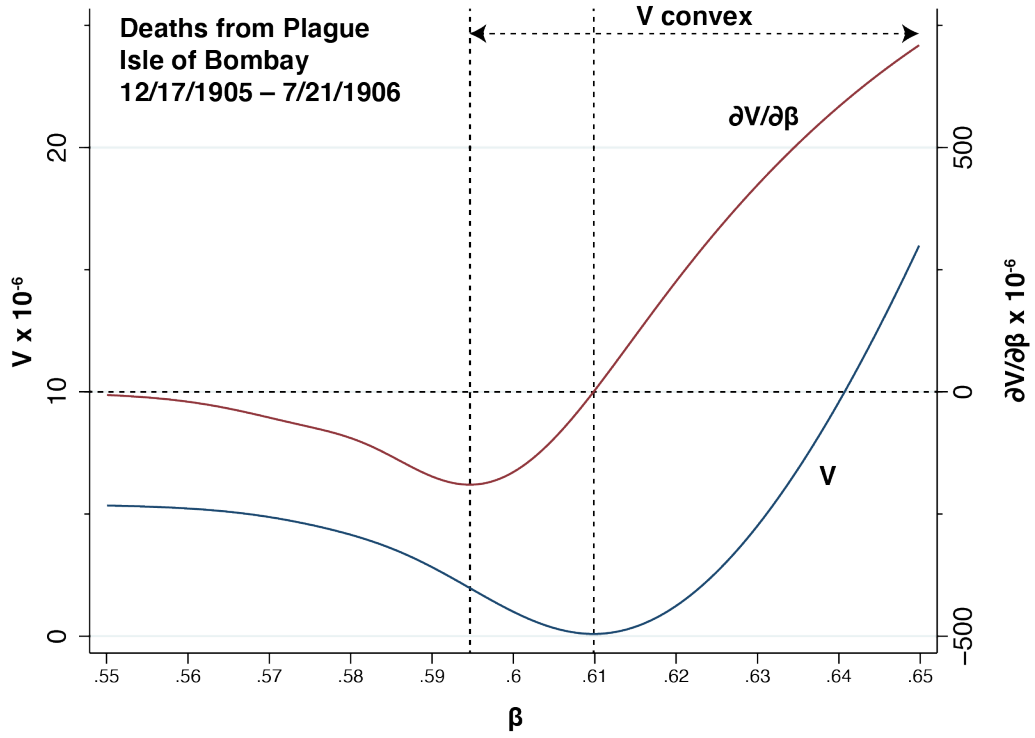

Figure H1. Least Squares Criterion V (Left Axis) and First Partial Derivative $\partial V/\partial \beta$ (Right Axis) as Functions of the Parameter β. Deaths from Plague, Isle of Bombay, 1905–1906.

Based upon the KM data on deaths from plague in the Isle of Bombay during 1905–1906, as shown in Figure 2 of the main text, Figure H1 above plots the least squares criterion V and the gradient $\partial V/\partial\beta$ as functions of the parameter β. As in corresponding Figure 4A in the main text, where the remaining parameters α, i_0 and N were held constant at their optimum values.

Figure H1 shows that the criterion function V is minimized at $\beta = 0.610$, as reported in the main text in connection with Figure 6, at which point the first partial derivative $\partial V/\partial\beta$ equals 0. The function $V(\beta)$ has an inflection point at $\beta = 0.595$, where the partial derivative $\partial V/\partial\beta$ reaches a minimum. In the interval $\beta > 0.595$, the function $V(\beta)$ is convex.

Declarations

Sole Authorship

JEH is the sole author of this work. He alone is responsible for the conceptualization of the work, the analysis of the data, the drafting of the manuscript, and the construction of the graphics. He warrants that it is entirely his original work. No requests for permission to use copyrighted material are required.

Acknowledgments

The opinions expressed here are solely those of the author and do not necessarily reflect those of the Massachusetts Institute of Technology, Eisner Health, or any other organization or individual. I gratefully acknowledge the helpful comments of the following individuals on Version 1 of this manuscript: Julio R. Banga (Consejo Superior de Investigaciones Científicas, Spain), Keith Godfrey (University of Warwick, UK), Weiguo Han (University Corporation for Atmospheric Research, USA), Şiran Keske (Koç University School of Medicine, Turkey), Eduardo Massad (University of Sao Paolo, Brazil), Thomas Michelitsch (Sorbonne Université, France), Dimiter Prodanov (IMEC, Belgium), Emerson Sadurni (Benemérita Universidad Autónoma de Puebla, Mexico), Reinhard Schlickeiser (Ruhr-Universität Bochum, Germany), Boris Schmid (University of Oslo, Norway), and James A. Yorke (University of Maryland, USA).

Availability of Data and Programs

The data and programs used in this work are publicly available on the Open Science Framework at https://osf.io/3b4hv/.

Human Subjects Declaration

This study relies exclusively on publicly available data that contain no individual identifiers. Links to the data sources are provided in the manuscript references.

Competing Interests Declaration

The author has no competing interests.

Funding Statement

This study did not receive any funding.

References

1. Capistran, M.A., M.A. Moreles, and B. Lara, *Parameter estimation of some epidemic models. The case of recurrent epidemics caused by respiratory syncytial virus.* Bull Math Biol, 2009. **71**(8): p. 1890-901.
2. Pollicott, M., H. Wang, and H.H. Weiss, *Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem.* J Biol Dyn, 2012. **6**: p. 509-23.
3. Marinov, T.T., et al., *Inverse problem for coefficient identification in SIR epidemic models.* Computers and Mathematics with Applications, 2014. **67**(12): p. 2218-2227.
4. Comunian, A., R. Gaburro, and M. Giudici, *Inversion of a SIR-based model: A critical analysis about the application to COVID-19 epidemic.* Physica D, 2020. **413**: p. 132674.
5. Kermack, W.O. and A.G. McKendrick, *A contribution to the mathematical theory of epidemics.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1927. **115**: p. 700-721.
6. Balkew, T.M., *The SIR Model When S(t) is a Multi-Exponential Function*. 2010, https://dc.etsu.edu/etd/1747/: East Tennessee State University, School of Graduate Studies, Electronic Theses and Dissertations, Paper 1747, December.
7. Shabbir, G., H. Khan, and M.A. Sadiq, *A note on Exact solution of SIR and SIS epidemic models*. 2010, https://arxiv.org/abs/1012.5035: ArXiv, Submitted December 22.
8. Harko, T., F.S.N. Lobo, and M.K. Mak, *Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates.* Applied Mathematics and Computation, 2014. **236**(1 Jun 2014): p. 184-194.
9. Kröger, M. and R. Schlickeiser, *Analytical solution of the SIR-model for the temporal evolution of epidemics. Part A: time-independent reproduction factor.* Journal of Physics A: Mathematical and Theoretical, 2020. **53**(505601): p. 1-38.
10. Sadurni, E. and G. Luna-Acosta, *Exactly solvable SIR models, their extensions and their application to sensitive pandemic forecasting.* Nonlinear Dyn, 2021. **103**(3): p. 2955-2971.
11. Harris, J.E., *The Coronavirus Epidemic Curve is Already Flattening in New York City*. 2020, https://www.nber.org/papers/w26917: National Bureau of Economic Research Working Paper No. 26917, Updated April 6, 2020.

12. Postnikov, E.B., *Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions?* Chaos, Solitons & Fractals, 2020. **135**(109841).
13. Magal, P. and G. Webb, *The parameter identification problem for SIR epidemic models: identifying unreported cases.* J Math Biol, 2018. **77**(6-7): p. 1629-1648.
14. Schmitt, F.G., *An algorithm for the direct estimation of the parameters of the SIR epidemic model from the I(t) dynamics.* The European Physical Journal Plus, 2021. **137**(57): p. 1-16.
15. Ma, J. and D.J. Earn, *Generality of the final size formula for an epidemic of a newly invading infectious disease.* Bull Math Biol, 2006. **68**(3): p. 679-702.
16. Arino, J., et al., *A final size relation for epidemic models.* Math Biosci Eng, 2007. **4**(2): p. 159-75.
17. Andreasen, V., *The final size of an epidemic and its relation to the basic reproduction number.* Bull Math Biol, 2011. **73**(10): p. 2305-21.
18. Mkhatshwa, T. and A. Mummert, *Modeling Super-spreading Events for Infectious Diseases: Case Study SARS*. 2010, https://arxiv.org/abs/1007.0908: arXiv, updated October 15.
19. Harris, J.E., *Data from the COVID-19 epidemic in Florida suggest that younger cohorts have been transmitting their infections to less socially mobile older adults.* Rev Econ Househ, 2020. **18**(4): p. 1019-1037.
20. Harris, J.E., *Los Angeles County SARS-CoV-2 Epidemic: Critical Role of Multi-generational Intra-household Transmission.* Journal of Bioeconomics, 2021. **23**: p. 55-83.
21. Rakshit, P., et al., *Modified SIR model for COVID-19 transmission dynamics: Simulation with case study of UK, US and India.* Results Phys, 2022. **40**: p. 105855.
22. Schlickeiser, R. and M. Krüger, *Key Epidemic Parameters of the SIRV-Model Determined from Past COVID-19 Mutant Waves*. 2023, https://www.preprints.org/manuscript/202303.0515/v1: Preprints.org, March 30.
23. Gopal, K., L.S. Lee, and H.-V. Seow, *Parameter Estimation of Compartmental Epidemiological Model Using Harmony Search Algorithm and Its Variants.* Applied Sciences, 2021. **11**(1138): p. 1+25.

24. Nguemdjo, U., et al., *Simulating the progression of the COVID-19 disease in Cameroon using SIR models.* PLoS One, 2020. **15**(8): p. e0237832.
25. Kucharski, A.J., et al., *Early dynamics of transmission and control of COVID-19: a mathematical modelling study.* Lancet Infect Dis, 2020. **20**(5): p. 553-558.
26. Rica, S. and G.A. Ruz, *Estimating SIR model parameters from data using differential evolution: an application with COVID-19 data*. 2020, https://ieeexplore.ieee.org/document/9277708: IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology (CIBCB)
27. Prodanov, D., *Analytical Parameter Estimation of the SIR Epidemic Model. Applications to the COVID-19 Pandemic.* Entropy, 2021. **23**(59): p. 1-20.
28. Capaldi, A., et al., *Parameter Estimation and Uncertainty Quantification for an Epidemic Model.* Mathematical Biosciences and Engineering, 2012. **9**(3): p. 553-579.
29. Turk, P.J., et al., *Modeling COVID-19 Latent Prevalence to Assess a Public Health Intervention at a State and Regional Scale: Retrospective Cohort Study.* JMIR Public Health Surveill, 2020. **6**(2): p. e19353.
30. Ahmetolan, S., et al., *What Can We Estimate From Fatality and Infectious Case Data Using the Susceptible-Infected-Removed (SIR) Model? A Case Study of Covid-19 Pandemic.* Front Med (Lausanne), 2020. **7**: p. 556366.
31. Li, M., J. Dushoff, and B.M. Bolker, *Fitting mechanistic epidemic models to data: A comparison of simple Markov chain Monte Carlo approaches.* Stat Methods Med Res, 2018. **27**(7): p. 1956-1967.
32. Su, L., et al., *Evaluation of the Secondary Transmission Pattern and Epidemic Prediction of COVID-19 in the Four Metropolitan Areas of China.* Front Med (Lausanne), 2020. **7**: p. 171.
33. Wangping, J., et al., *Extended SIR Prediction of the Epidemics Trend of COVID-19 in Italy and Compared With Hunan, China.* Front Med (Lausanne), 2020. **7**: p. 169.
34. Cauchemez, S. and N.M. Ferguson, *Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London.* J R Soc Interface, 2008. **5**(25): p. 885-97.
35. Gu, J. and G. Yin, *Bayesian SIR model with change points with application to the Omicron wave in Singapore.* Sci Rep, 2022. **12**(1): p. 20864.

36. Koepke, A.A., et al., *Predictive Modeling of Cholera Outbreaks in Bangladesh.* Annals of Applied Statistics, 2016. **10**(2): p. 575-593.
37. Cooper, I., A. Mondal, and C.G. Antonopoulos, *A SIR model assumption for the spread of COVID-19 in different communities.* Chaos Solitons Fractals, 2020. **139**: p. 110057.
38. Finkenstädt, B.F. and B.T. Grenfell, *Time Series Modelling of Childhood Diseases: A Dynamical Systems Approach.* Journal of the Royal Statistical Society. Series C (Applied Statistics), 2000. **49**(2): p. 187-205.
39. Wanduku, D. and C. Rahul, *Complete maximum likelihood estimation for SEIR epidemic models: theoretical development*. 2019, https://arxiv.org/abs/1907.10679: arXiv, Updated July 29.
40. Diekmann, O., et al., *The discrete-time Kermack-McKendrick model: A versatile and computationally attractive framework for modeling epidemics.* Proc Natl Acad Sci U S A, 2021. **118**(39).
41. Cintron-Arias, A., et al., *The estimation of the effective reproductive number from disease outbreak data.* Math Biosci Eng, 2009. **6**(2): p. 261-82.
42. Hooker, G., et al., *Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario.* J R Soc Interface, 2011. **8**(60): p. 961-74.
43. Chowell, G., et al., *Modelling the transmission dynamics of acute haemorrhagic conjunctivitis: application to the 2003 outbreak in Mexico.* Stat Med, 2006. **25**(11): p. 1840-57.
44. Gabor, A. and J.R. Banga, *Robust and efficient parameter estimation in dynamic models of biological systems.* BMC Syst Biol, 2015. **9**: p. 74.
45. Schroeder, C., *Practical course on computing derivatives in code*, in *Proceedings of SIGGRAPH '19 Courses, July 28 - August 01, 2019, Los Angeles, CA, USA*. 2019, ACM: https://dl.acm.org/doi/10.1145/3305366.3328073.
46. Verbeke, J. and R. Cools, *The Newton-Raphson method.* International Journal of Mathematical Education in Science and Technology, 1995. **26**(2): p. 177-193.
47. Hartley, H.O., *The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares.* Technometrics, 1961. **3**(2): p. 269-280.
48. Gallant, A.R., *Nonlinear Regression.* The American Statistician, 1975. **29**(2 (May)): p. 73-81.

49. Kermack, W.O. and A.G. McKendrick, *Contributions to the mathematical theory of epidemics, part II.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1932. **138**(834): p. 55-83.
50. Kermack, W.O. and A.G. McKendrick, *Contributions to the mathematical theory of epidemics, part III.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1933. **141**(843): p. 55-83.
51. Anderson, R.M., *Discussion: the Kermack-McKendrick epidemic threshold theorem.* Bull Math Biol, 1991. **53**(1-2): p. 3-32.
52. Brauer, F., *Mathematical epidemiology: Past, present, and future.* Infect Dis Model, 2017. **2**(2): p. 113-127.
53. Brown, K.S. and J.P. Sethna, *Statistical mechanical approaches to models with many poorly known parameters.* Physical Review E, 2003. **68**(021904): p. 1-9.
54. Evans, N.D., et al., *The structural identifiability of the susceptible infected recovered model with seasonal forcing.* Math Biosci, 2005. **194**(2): p. 175-97.
55. Gutenkunst, R.N., et al., *Universally Sloppy Parameter Sensitivities in Systems Biology Models.* PLOS Computational Biology, 2007. **3**(10 (e189)): p. 1871-1878.
56. Massonis, G., J.R. Banga, and A.F. Villaverde, *Structural identifiability and observability of compartmental models of the COVID-19 pandemic.* Annu Rev Control, 2021. **51**: p. 441-459.
57. Smith, D.A. and L.C. Moore, *Project 2: The SIR Model of the Spread of Disease - Euler's Method for Systems*, in *Calculus: Modeling and Application, 2nd Edition (Interactive Textbook)*. 2010, Mathematical Association of America: http://calculuscourse.maa.org/sample/Chapter5/Projects/SIR%20Model/SIR4.html.
58. Guldberg, C.M. and P. Waage, *Studies Concerning Affiinity: Presentation to Academy of Sciences in Christiania (Oslo). Translated by Henry I. Abrash. Originally published by Forhandlinger: Videnskabs-Selskabet i Christiana, 1864.* Journal of Chemical Education, 1986. **63**(12): p. 1044-1047.
59. Hethcote, H.W., *Qualitative analyses of communicable disease models.* Mathematical Biosciences, 1976. **28**(3-4): p. 335-356.
60. Weiss, H.H., *The SIR model and the Foundations of Public Health.* Materials Matemàtics, 2013. **3**: p. 1-17.

61. van den Driessche, P. and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission.* Math Biosci, 2002. **180**: p. 29-48.
62. Keeling, M.J., et al., *Fitting to the UK COVID-19 outbreak, short-term forecasts and estimating the reproductive number.* Stat Methods Med Res, 2022. **31**(9): p. 1716-1737.
63. Becker, N.G. and T. Britton, *Statistical studies of infectious disease incidence.* J R Statist Soc B, 1999. **61**(2): p. 287-307.
64. Lange, A., *Reconstruction of disease transmission rates: Applications to measles, dengue, and influenza.* J Theor Biol, 2016. **400**: p. 138-53.
65. Transtrum, M.K. and J.P. Sethna, *Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization*. 2012, https://arxiv.org/abs/1201.5885: arXiv.org, issued January 27.
66. Gabor, A., A.F. Villaverde, and J.R. Banga, *Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems.* BMC Syst Biol, 2017. **11**(1): p. 54.
67. Oehlert, G.W., *A Note on the Delta Method.* The American Statistician, 1992. **46**(1): p. 27–29.
68. Dempster, A.P., N.M. Laird, and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society, Ser. B, 1977. **39**: p. 1-22.
69. Qasmieh, S.A., et al., *The Importance of Incorporating At-Home Testing Into SARS-CoV-2 Point Prevalence Estimates: Findings From a US National Cohort, February 2022.* JMIR Public Health Surveill, 2022. **8**(12): p. e38196.
70. Abadie, A., et al., *Epidemic Modeling and Estimation*. 2020, https://idss.mit.edu/wp-content/uploads/2020/04/04.28.2020Epidemic_Modeling_A_Memo.pdf: Massachusetts Institute of Technology, Institute for Data Systems and Society (IDSS), April 28.
71. Roda, W.C., et al., *Why is it difficult to accurately predict the COVID-19 epidemic?* Infect Dis Model, 2020. **5**: p. 271-281.
72. Trejo, I. and N.W. Hengartner, *A modified Susceptible-Infected-Recovered model for observed under-reported incidence data.* PLoS One, 2022. **17**(2): p. e0263047.
73. Brauer, F., *Epidemic models with heterogeneous mixing and treatment.* Bull Math Biol, 2008. **70**(7): p. 1869-85.

74. Kong, L., et al., *Modeling Heterogeneity in Direct Infectious Disease Transmission in a Compartmental Model.* Int J Environ Res Public Health, 2016. **13**(3).
75. David, J.F., *Epidemic models with heterogeneous mixing and indirect transmission.* J Biol Dyn, 2018. **12**(1): p. 375-399.
76. Duan, M. and Z. Jin, *The heterogeneous mixing model of COVID-19 with interventions.* J Theor Biol, 2022. **553**: p. 111258.
77. Harris, J.E., *Concentric regulatory zones failed to halt surging COVID-19: Brooklyn 2020.* Front Public Health, 2022. **10**: p. 970363.
78. Hill, A.N., J.W. Glasser, and Z. Feng, *Implications for infectious disease models of heterogeneous mixing on control thresholds.* J Math Biol, 2023. **86**(4): p. 53.
79. Harris, J.E., *Geospatial Analysis of a COVID-19 Outbreak at the University of Wisconsin - Madison: Potential Role of a Cluster of Local Bars.* Epidemiol Infect, 2022: p. 1-31.
80. New York Department of Health and Mental Hygiene, *NYC Coronavirus Disease 2019 (COVID-19) Data*. 2022, https://github.com/nychealth/coronavirus-data: Accessed July 3.
81. StataCorp, *Stata Statistical Software: Release 17*. 2021, College Station, TX: StataCorp LLC.
82. StataCorp, *Stata Programming Reference Manual Release 17*. 2021, https://www.stata.com/manuals/p.pdf: College Station, Texas: Stata Press.
83. StataCorp, *Mata Reference Manual Release 17*. 2021, https://www.stata.com/manuals/m.pdf: College Station, Texas: Stata Press.
84. Mander, A., *SURFACE: Stata module to draw a 3D wireform surface plot*. 2019, https://ideas.repec.org/c/boc/bocode/s448501.html: Boston College Department of Economics, Statistical Software Components S448501, initial version 2005, revised 24 Jun 2019.
85. Keske, S., et al., *Duration of infectious shedding of SARS-CoV-2 Omicron variant and its relation with symptoms.* Clin Microbiol Infect, 2022.
86. Timmins, C. and W. Schlenker, *Reduced-Form Versus Structural Modeling in Environmental and Resource Economics.* Annual Review of Resource Economics, 2009. **1**: p. 351-380.
87. Hespanha, J.P., et al., *Forecasting COVID-19 cases based on a parameter-varying stochastic SIR model.* Annual Reviews in Control, 2021. **51**: p. 460-476.

88. Monecke, S., H. Monecke, and J. Monecke, *Modelling the black death. A historical case study and implications for the epidemiology of bubonic plague.* Int J Med Microbiol, 2009. **299**(8): p. 582-93.

89. Lewnard, J.A. and J.P. Townsend, *Climatic and evolutionary drivers of phase shifts in the plague epidemics of colonial India.* Proc Natl Acad Sci U S A, 2016. **113**(51): p. 14601-14608.

90. Kidambi, P., *The Making of an Indian Metropolis: Colonial Governance and Public Culture in Bombay, 1890-1920*. 2016, London: Routledge.

91. Frazier, M.W., *Tale of two Mumbai epidemics: Striking parallels between 1896 bubonic plague and 2020 Covid outbreak*. 2020, https://scroll.in/article/976706/tale-of-two-mumbai-epidemics-striking-parallels-between-1896-bubonic-plague-and-2020-covid-outbreak: Scroll.in, November 13.

92. Kugeler, K.J., et al., *Epidemiology of human plague in the United States, 1900-2012.* Emerg Infect Dis, 2015. **21**(1): p. 16-22.

93. Drancourt, M., L. Houhamdi, and D. Raoult, *Yersinia pestis as a telluric, human ectoparasite-borne organism.* Lancet Infect Dis, 2006. **6**(4): p. 234-41.

94. Gani, R. and S. Leach, *Epidemiologic determinants for modeling pneumonic plague outbreaks.* Emerg Infect Dis, 2004. **10**(4): p. 608-14.

95. Dean, K.R., et al., *Human ectoparasites and the spread of plague in Europe during the Second Pandemic.* Proc Natl Acad Sci U S A, 2018. **115**(6): p. 1304-1309.

96. Whittles, L.K. and X. Didelot, *Epidemiological analysis of the Eyam plague outbreak of 1665-1666.* Proc Biol Sci, 2016. **283**(1830).

97. Didelot, X., L.K. Whittles, and I. Hall, *Model-based analysis of an outbreak of bubonic plague in Cairo in 1801.* J R Soc Interface, 2017. **14**(131).

98. Lien-Teh, W., *A Treatise on Pneumonic Plague*. 1926, https://wellcomecollection.org/works/drfupc3x/items: Geneva: League of Nations Health Organization.

99. Massad, E., et al., *The Eyam plague revisited: did the village isolation change transmission from fleas to pulmonary?* Med Hypotheses, 2004. **63**(5): p. 911-5.

100. Qasmieh, S.A., et al., *The prevalence of SARS-CoV-2 infection and other public health outcomes during the BA.2/BA.2.12.1 surge, New York City, April-May 2022*. 2022, https://www.medrxiv.org/content/10.1101/2022.05.25.22275603v3: medRxiv, July 18.

101. Qasmieh, S.A., et al., *Estimating the period prevalence of SARS-CoV-2 infection during the Omicron (BA.1) surge in New York City (NYC), January 1-March 16, 2022*. 2022, https://www.medrxiv.org/content/10.1101/2022.04.23.22274214v1: medRxiv, April 27.

102. Dejnirattisai, W., et al., *SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses.* Cell, 2022. **185**(3): p. 467-484 e15.

103. Moss, P., *The T cell immune response against SARS-CoV-2.* Nat Immunol, 2022. **23**(2): p. 186-193.

104. Harris, J.E., *Mobility was a significant determinant of reported COVID-19 incidence during the Omicron Surge in the most populous U.S. Counties.* BMC Infect Dis, 2022. **22**(1): p. 691.

105. FIBIwiki, *Bombay (City)*. 2022, https://wiki.fibis.org/w/Bombay_(City): Fibis.org. Last updated November 14, 2022. Accessed February 25, 2023.

106. Andrianaivoarimanana, V., et al., *Short- and long-term humoral immune response against Yersinia pestis in plague patients, Madagascar.* BMC Infect Dis, 2020. **20**(1): p. 822.

107. Munoz-Fernandez, G.A., J.M. Seoane, and J.B. Seoane-Sepulveda, *A SIR-type model describing the successive waves of COVID-19.* Chaos Solitons Fractals, 2021. **144**: p. 110682.

108. Bestehorn, M., et al., *Simple model of epidemic dynamics with memory effects.* Physical Review E, 2022. **105**(024025): p. 024025.

109. New York City Department of Health and Mental Hygiene, *variant-epi-data*. 2023, https://github.com/nychealth/coronavirus-data/tree/master/variants: Accessed April 21.

110. Cox, N.J., *PCHIPOLATE: Stata module for piecewise cubic Hermite interpolation*. 2012, https://ideas.repec.org/c/boc/bocode/s457561.html: Statistical Software Components S457561, Boston College Department of Economics. Accessed April 21.

111. U.S. Centers for Disease Control and Prevention, *Ending Isolation and Precautions for People with COVID-19: Interim Guidance*. 2022, https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html: Updated January 14.

112. Altarawneh, H.N., et al., *Effects of Previous Infection and Vaccination on Symptomatic Omicron Infections.* N Engl J Med, 2022. **387**(1): p. 21-34.

113. Los Angeles County Department of Public Health, *COVID-19 Surveillance Dashboard*. 2022, https://lacdph.shinyapps.io/covid19_surveillance_dashboard/: Accessed July 3.