Supplemental Materials

Additional methodology details

We consider the following observation model for infections and confirmed cases. Each individual $n \in \{1, \ldots, N\}$ who eventually becomes infected is confirmed to be infected on day $C_n \in \{1, \ldots, T\}$, and we assume that there were no infections prior to the initial time, denoted 1. The time of a confirmed case is stochastically delayed from the true date of infection $I_n \in \{1, \ldots, T\}$,

$$D_n \sim \text{Categorical}(\theta) \tag{4}$$

$$C_n = I_n + D_n \,, \tag{5}$$

where D_n is a random number of days sampled from a discrete distribution with probability vector $\theta = (\theta_0, \ldots, \theta_P)$. We assume that θ is available as external knowledge from an auxiliary data source and consider it fixed throughout. For example, COVID-19 line lists track the number of days between symptom development and reported case in individual cases can be used to form an estimate of the distribution θ [1, 2]. While the methods considered condition on a specific θ, uncertainty in θ can be explored by a sensitivity analysis or averaging over a distribution of θ estimates.

Daily incidence is a summary of all cases from a given day. We write confirmed case incidence on day t as $Y = (Y_1, \ldots, Y_T)$, where $Y_t = \sum_n^N \mathbb{1}(C_n = t)$ and $\mathbb{1}(C_n = t)$ takes the value one when C_n is equal to t and zero otherwise. Similarly, the unobserved infection incidence curve is defined as $X = (X_1, \ldots, X_T)$, where $X_t = \sum_n^N \mathbb{1}(I_n = t)$. The infection incidence curve is the target of our estimator. We depict an example of a delay distribution θ and the resulting infection incidence and observed incidence pair in eFigure 1a.

Observed case likelihood

The likelihood corresponding to the observed time series, $\boldsymbol{y}$ is a function of the convolution of the infection incidence $X = X_1, \ldots, X_T$ and the delay distribution θ. To see this, we view each marginal Y_t as the sum of N Poisson processes. Consider the infection date I_n for individual n. We can view the value C_n as the realization of a Poisson process with intensity function θ offset to start at the value of I_n, conditioned on a single realization.

Take all X_s individuals infected on day s. Due to the additivity of Poisson processes, the number of cases reported on day t that originated on day s will follow a Poisson distribution, which we denote

$$Y(s, t) \sim \text{Pois}(X_s \cdot \theta_{t-s}) \,, \tag{6}$$

where we define $\theta_s = 0$ for $s < 0$ and $s > P$. The total number of reported cases on day t is a sum of $Y(s, t)$ over its first argument up to day t, which is again Poisson

$$Y_t \sim \sum_s Y(s, t) \sim \text{Pois}\left(\sum_s X_s \cdot \theta_{t-s}\right) \,, \tag{7}$$

where the mean is the familiar convolution operator. Conditioning on the total number infected, N, the vector $Y = (Y_1, \ldots, Y_T)$ is jointly multinomial

$$Y_1, \ldots, Y_T \sim \text{Mult}(p; N) \tag{8}$$

$$p = \left(\sum_s X_s \cdot \theta_{1-s}, \ldots, \sum_s X_s \cdot \theta_{T-s}\right) \,. \tag{9}$$

The expected number of reported cases, $E[Y] = (E[Y_1], \ldots, E[Y_T])$ is related to the underlying infection time series $X = (X_1, \ldots, X_T)$ by the convolution

$$E[Y] = X * \theta \,. \tag{10}$$

This is equivalent to a linear mapping of X, $E[Y] = P_\theta X$ for the matrix P_θ constructed to correspond to the discrete convolution with vector θ

$$P_\theta = \begin{pmatrix} \theta_0 & & & & \\ \theta_1 & \theta_0 & & & \\ \theta_2 & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \theta_T & \ldots & \theta_2 & \theta_1 & \theta_0 \end{pmatrix} . \tag{11}$$

All unidentified entries of P_θ are 0. Intuitively, recovering X from Y involves the *inverse* of this mapping. However, a popular reconstruction approach — which we term the "re-convolution" estimator — does not correspond to the inverse of this mapping, but actually to another application of the same convolution.

"Re-convolution" incidence reconstruction

An intuitive approach to infection incidence reconstruction is to stochastically "undo" the results of the delay random variable, D_n. One type of estimator for this is to simply sample a new $D_n \sim \text{Cat}(\theta)$ from the delay distribution and subtract this new value from the observed case date

$$D_n \sim \text{Cat}(\theta) \,, \tag{12}$$

$$\hat{I}_n = c_n - D_n \,, \tag{13}$$

$$\hat{X}_t = \sum_n \mathbb{1}(\hat{I}_n = t) \,. \tag{14}$$

A related approach is to "undo" the convolution of θ with I by running the convolution backward over the observed case counts, resulting in the estimator

$$\hat{X}_s = \sum_t Y_t \cdot \theta_{t-s} \,, \tag{15}$$

$$\hat{X} = (\hat{X}_1, \ldots, \hat{X}_T) = Y \,\tilde{*}\, \theta \,, \tag{16}$$

where $\tilde{*}$ indicates reversing the direction of the convolution operator. Intuitively, this operation distributes observed cases backward in time according to the delay distribution. This has become a popular approach to cope with delayed observations [3, 4, 5, 6, 7].
This reverse convolution corresponds to multiplication by the *transpose* of P_θ

$$\hat{X} = P_\theta^\intercal Y \,, \tag{17}$$

which effectively applies *another* convolution to the already convolved time series Y, further smoothing the already smoothed observation. We term this estimator the "re-convolution" estimator.
Given that $E[Y] = P_\theta X$, a conceptually straightforward method for computing X is to simply invert P_θ, noting that $X = P_\theta^{-1} E[Y]$. Because we only have access to Y, and not $E[Y]$, one might derive a plug-in estimator, substituting the observation Y for $E[Y]$

$$\hat{X} = P_\theta^{-1} Y \,. \tag{18}$$

This estimator, however, is unstable in the presence of noise in the daily case counts Y, rendering it impractical for applied analysis. Intuitively, this inverted convolution estimator is similar to fitting a linear regression model where the number of observations is equal to the number of covariates (i.e., $n = p$) — an ill-posed inverse problem [8]. We refer to this estimator as the "unstable" deconvolution estimator.

Using the "re-convolution" estimator to impute the incidence curve results in an even smoother imputed curve than the already smoothed reported case incidence curve. Intuitively that makes sense — the "re-convolution" method convolves the already convolved reported case curve with the delay distribution once more, again smoothing (but backward in time).

Using the direct inverse of P_θ is only practical when we have access to the true expected value $E[Y]$. The "unstable" estimator directly applies P_θ^{-1} to the noisy observed time series, but results in highly unstable estimates — we see the estimate vary between -10^{20} and 10^{20} cases. We note that the instability can be even worse in boundary regions, where incidence may be determined by only a few days of observed cases. This instability motivates our development of a practical deconvolution estimator.

Model-based deconvolution estimators

As stated in the *Methods* section, model-based estimators start with a likelihood model for observed case data, conditioned on the underlying incidence curve. Concretely, model-based methods define an objective

$$L(\beta; Y_1, \ldots, Y_T) = \underbrace{-\ln Pr(Y_1, \ldots, Y_T \mid \beta)}_{\text{neg. log likelihood}} + \underbrace{\lambda \cdot r(\beta)}_{\text{regularization}} , \tag{19}$$

where the form of the log-likelihood is specified by a set of model parameters, β, a probabilistic model, and the regularization penalty enforces smoothness in the latent incidence curve. Back projection and back calculation methods have examined the use of Poisson and multinomial likelihoods [9, 10].

For the Robust Incidence Deconvolution Estimator (RIDE), we use a Poisson likelihood and enforce smoothness in the latent time series by parameterizing X with a spline model

$$X_t = \exp\left(\beta^\intercal B(t)\right), \text{ for } t = 1, \ldots, T, \tag{20}$$

where $\beta \in \mathbb{R}^Q$ and $B(t)$ is the Q-dimensional spline basis value at time t. The conditional distribution of Y_t is Poisson, with the mean parameter defined by the convolution of X with the delay distribution, $Y_t \sim \text{Pois}(\lambda_t)$ and $\lambda_t = (P_\theta X)_t$.

We also note that the model-based method is closely related to a set of techniques known as empirical Bayes. In fact, the empirical Bayesian method g-modeling was developed to de-noise (or deconvolve) noisy observed data [11]. The basic idea behind g-modeling is to fit a flexible prior distribution over the unobserved quantity using marginal maximum likelihood. The per-subject unobserved values are the infection times I_n for $n = 1, \ldots, N$, and the noisy observations are confirmed case dates c_n. The g-modeling approach parameterizes the prior (i.e., models the g-function), and numerically optimizes this parameter by maximizing the marginal likelihood (i.e., the f-function). Specifically we define a prior over I_n parameterized by β

$$g_\beta(s) \equiv Pr(I_n = s) \tag{21}$$

and fit β values by maximizing the marginal likelihood of the observed data under the noise

model given by θ

$$\hat{\beta} = \arg\max_\beta \sum_n \ln f(c_n) + \lambda r(\beta)\,, \tag{22}$$

$$= \arg\max_\beta \sum_n \ln \int g_\beta(s) p(c_n \mid s) ds + \lambda r(\beta)\,. \tag{23}$$

Given this estimate, we can compute the de-noised incidence curve via Bayes rule. As a concrete example, $g_\beta(s)$ can be parameterized by a set of basis functions

$$g_\beta(s) \propto \exp(\beta^\intercal B(s)) \tag{24}$$

$$B(s) = (B_1(s), \ldots, B_P(s)) \qquad \text{e.g., } P\text{-dimensional spline basis.} \tag{25}$$

Here, the distribution $g_\beta(s)$ is determined by a log-linear function of splines, and is normalized. This empirical Bayes approach leads to a slightly different likelihood — a multinomial. While performance can be similar, we found the Poisson likelihood model easier to fit.

Regularization

To form estimates, we must choose a set of hyperparameters, including the number of spline basis functions (i.e., the degree of freedom (DoF) parameter, which controls how complex the sample paths can be) and the regularization strength λ.

Additionally, to handle censoring, we treat data to the right of the observation window as missing, and impute plausible sample paths, averaging over their uncertainty [12]. Because the smoothing of the stochastic delay typically smooths observations and induces a temporal autocorrelation among reported cases, we extrapolate the report curve with a simple random walk. We first apply an Anscombe transform [13] (i.e., a variance stabilizing transform) and use the empirical single-lag autocorrelation to simulate a set of random walk imputations. We use a first order extrapolation procedure, which offers significant improvements over a zeroth order extrapolation or zero imputation. Nevertheless, it can be sensitive to outliers and produce under-estimates when there is under-ascertainment. More complicated now-cast methods like that of [14] can correct for under-ascertainment when recording delay data are available.

We use a data-driven selection procedure:

- Select spline DoF:
  - Set the default λ to be $1/\sum_t Y_t$
  - Compute the $\hat{\beta}$ estimate for a grid of DoF
  - Select the DoF parameter with the lowest AIC (or BIC)
- Select λ by data splitting — repeat the following four times (by default) and average
  - Split the data into (by default) 25% validation and 75% training by randomly thinning each observed count
  - Compute the $\hat{\beta}$ estimate for a grid of candidate λ values (with selected DoF value)
  - Select the largest (i.e., smoothest) λ within (by default) 2% of the best observed held out log likelihood

The rest of the procedure is as follows: for each imputed time-series, form a set of samples

- Compute the $\hat{\beta}$ estimate given the selected λ and DoF parameters
- Sample β from a Laplace approximation formed at the $\hat{\beta}$ estimate
- Compute the sample path $X_{1:T}$ for this β sample

This procedure yields a collection of $X_{1:T}$ samples, which we use to form the incidence estimate.

Uncertainty estimation

We construct uncertainty intervals over the estimate of X_t over a set of Laplace approximations, one for each of the auto-regression (AR) replicates. For each AR replicate, we compute $\hat{\beta}$ by minimizing the negative log likelihood function defined in Equation 3. To approximate uncertainty around the β parameters, we use a Laplace approximation, $\beta \sim \mathcal{N}(\hat{\beta}, \hat{\Sigma})$, where $\hat{\Sigma} \equiv H_{\hat{\beta}}^{-1}$ is the Hessian of the negative log likelihood evaluated at the estimate $\hat{\beta}$. Note that this Hessian includes the regularization term, which can be interpreted as a prior over β. For each AR replicate, we generate L samples $\{\beta^{(\ell)}\}_{\ell=1}^{L}$ from this normal approximation. For each sample ℓ we compute the implied $X_{1:T}^{(\ell)}$ from the spline model, and store all samples over all AR replicates. Quantiles of these combined samples are then used to compute credible intervals.

Finally, we find that this estimation procedure tends to find overly narrow coverage intervals, likely due to overly smooth model assumptions. Therefore, when computing coverage intervals, we inflate the percentiles by a square root factor, $u_{adjusted} = u + \sqrt{u}$ and $l_{adjusted} = l - \sqrt{l}$. This creates an overdispersion effect, which is often seen in infection time series but is missing from a Poisson model. Empirically, we find that this post-hoc adjustment works well. In fact, the coverage frequencies of both the re-convolution and back projection estimators could benefit from a similar adjustment.

eFigure 8 shows the average coverage over all experimental settings.

Additional synthetic data results

The synthetic models are specified as follows. All of the shapes are normalized to have density 1 and then re-scaled to have 5,000 infections distributed over 100 time steps. All delay distributions are Gamma(10,1). For the well-specified model, each case on the incidence curve is randomly propagated forward s days according to the delay distribution. For the noisy (non-well-specified) case, every sixth and seventh day a uniform random number between 0.3 and 0.5 is drawn and that proportion of cases are recorded two days later (e.g. cases from day six move to day eight). This is done to approximate reporting delays for testing and death records.

Symmetric. Incidence has a Gaussian shape,

$$X_t \propto \exp\left\{-0.0025\,(t-50)^2\right\},$$

for $t = 0, \ldots, 100$.

Symmetric Censored. Incidence is the same as Symmetric, except that the incidence curve and reporting stops at $t = 65$.

Slow Decay. Incidence increases quickly then declines slowly,

$$X_t \propto \begin{cases} \exp\left\{-0.05(8-t)^2\right\} & 0 \le t \le 8, \\ \exp\left\{-0.001(t-8)^2\right\} & t > 8, \end{cases}$$

for $t = 0, \ldots, 88$.

Pathological. Incidence increases quickly then stops entirely,

$$X_t \propto \begin{cases} \exp\left\{t/4\right\} & 0 \le t \le 29, \\ 0 & t > 29, \end{cases}$$

for $t = 0, \ldots, 79$.

Double Peak. Incidence has two symmetric peaks in succession,

$$X_t \propto 0.55 \exp\left\{-\frac{1}{2*8^2}(t-30)^2\right\} + 0.35 \exp\left\{-\frac{1}{2*7^2}(t-60)^2\right\} + 0.1 \exp\left\{-\frac{1}{2*10^2}(t-70)^2\right\}$$

for $t = 0, \ldots, 100$.

Matern and Matern2. To obtain a time series with more direction changes, we randomly generate two curves from a Gaussian process model with a Matérn covariance function. The Matérn covariance has three parameters, ν controls the smoothness (i.e., when $\nu \to \infty$ samples become infinitely differentiable), σ controls the marginal variance of sample paths, and ℓ controls the lengthscale (i.e., larger ℓ implies more correlation between nearby points). Setting $\nu = 3/2$, the Matérn covariance function between observations at times s and t such that $d = |s - t|$ can be written

$$k(d) = \sigma^2 \left(1 + \frac{d}{\ell}\right) \exp\left(-\frac{d}{\ell}\right). \tag{26}$$

The latent infection time series can be sampled from a multivariate normal distribution

$$\ln X_{1:T} \sim \mathcal{N}(0, K) \tag{27}$$

where the Gram matrix K is defined such that $K_{s,t} = k(|X_s - X_t|)$. We generate two curves from different random seeds, which we term `matern` and `matern2` in the experiments.

Results

All synthetic curves and associated observations are depicted in eFigure 2a.
eFigure 1 shows RMSE results for each of the six curves, broken down by overall RMSE and RMSE on the most recent 20 days. eFigure 2 further presents average RMSE by curve type, noise setting, and observation window.
eFigures 3–7 show fits for each model under well-specified and mis-specified regimes, with the full time series and the truncated time series.
eFigures 8a and 8b plot coverage rates for the 95% interval for each of the estimators overall and for the most recent 20 days, respectively.

Comparison of Robust Inference Deconvolution Estimator hyperparameter choices

RIDE fits a deconvolution model using a cubic spline basis and second order regularization. Many different regularization methods and basis function types have been proposed in previous literature; see Table 1 in the main text.
First, we compare RIDE basis choice: cubic splines vs. step functions on a daily scale. Fits are shown in the left panel of eFigure 9 for Staten Island and Arizona. We note that the fits are similar, although the step function basis fit requires one to two more orders of magnitude more computing time than a cubic spline basis fit due to the larger number of parameters to be fitted.
Next, we compare the order of the regularization penalty: 0, 1, or 2. Fits are shown in eFigure 9. First and second order penalties produce similar fits, although first order fits usually have more derivative sign changes than second order fits. Zeroth order penalties are generally under-regularized for daily COVID-19 reporting data. The presence vs. absence of extrapolation is shown in eFigure 9. The presence of regularization prevents right tail overfitting in Arizona.

Other tunable hyperparameters, such as the regularization penalty and the degrees of freedom in the spline basis, are chosen through a hyperparameter sweep and evaluation on a validation set. The most regularizing parameter values that produce fits within 2% of the best likelihood value are selected in the "validation" method. This is the default selection method for λ, the regularization penalty. Default selection method for the cubic spline degrees of freedom is the Akaike Information Criterion (AIC), which is a less computationally intensive asymptotic approximation of held out likelihood. Comparison of hyperparameter selection methods are also shown in eFigure 9.

Comparison of back projection hyperparameter choices

We used the back projection of [10] for comparison as it is readily available for community usage. It was fit by the `backprojNP` method in the `surveillance` package version 1.18.0 in R version 4.0.0 with parameters iter.max=rep(250,2) and B=-1. We used case data from states/regions not used in the comparison to choose tunable parameters, k, which controls neighborhood size in the EM smoothing step, and epsilon, which controls stopping criteria. We varied k across values $k = 0, 2, 4, 8, 16, 32, 64, 128$, and epsilon across values $eps = 0.1, 0.001, 0.00001$.
Fits are shown for Brooklyn and Manhattan case data in eFigure 10. Error between reported daily cases and fitted incidence convolved with the delay distribution is shown in eTable 1.
We note that fitting error is minimized when $k = 0$ in all cases, although this leads to unstable incidence fits. To balance the trade off between fitting error and fit stability, we chose $k = 16$. As the choice of epsilon is relatively minimal for all values except $k = 0$, we kept the default of 0.005.
We also note that the choice of tunable parameters for this method involves much more manual selection than RIDE and seems to produce parameter values that are less robust to out of sample data.

Data

All case and death data for regions other than New York City is obtained from the New York Times database [`https://github.com/nytimes/covid-19-data`]. Hospitalization data are gathered from other sources in all regions.

Arizona. Hospitalization data are from the Arizona Department of Health [`https://www.azdhs.gov/preparedness/epidemiology-disease-control/infectious-disease-epidemiology/covid-19/dashboards/index.php`].

Ohio. Hospitalization data are from the Ohio Department of Health [`https://coronavirus.ohio.gov/static/COVIDSummaryData.csv`].

Texas. Areas are defined as follows: Greater Houston area is defined as Austin, Brazoria, Fort Bend, Galveston, Harris, Montgomery, and Waller counties; the Dallas/Fort Worth areas is Collin, Dallas, Denton, Ellis, Hood, Hunt, Johnson, Kaufman, Parker, Rockwall, Somervell, Tarrant, and Wise counties; the Austin area is Travis, Hays, Williamson, Bastrop, and Caldwell counties.

New York. All New York City data, including cases, deaths, and hospitalizations, is obtained from the New York City Department of Health [`https://raw.githubusercontent.com/nychealth/coronavirus-data/master/boro/boroughs-case-hosp-death.csv`]. State level case and death data are from the New York Times.

k	epsilon	MSE			
		Brooklyn	Manhattan	Virginia	Ohio
0	1E-5	9,844.4	3,306.9	46,255.7	676,201.1
	1E-3	9,844.4	3,306.9	46,255.7	676,201.1
	1E-1	12,143.5	3,849.0	66,102.4	725,340.4
2	1E-5	11,369.7	3,652.5	54,849.9	703,651.2
	1E-3	11,393.5	3,658.1	55,106.3	704,021.3
	1E-1	12,173.2	3,853.8	66,275.0	725,872.0
4	1E-5	11,574.9	3,705.3	57,682.1	710,369.2
	1E-3	11,589.1	3,710.0	57,954.9	710,542.5
	1E-1	12,200.6	3,858.0	66,442.2	726,471.6
8	1E-5	11,748.8	3,759.0	60,872.7	716,864.2
	1E-3	11,756.1	3,762.6	61,043.8	716,957.5
	1E-1	12,250.6	3,865.0	66,762.9	727,658.9
16	1E-5	11,893.6	3,807.4	63,895.8	722,635.4
	1E-3	11,897.6	3,809.2	63,962.5	722,713.1
	1E-1	12,339.1	3,876.5	67,341.5	729,777.1
32	1E-5	12,080.0	3,851.8	66,446.6	727,986.6
	1E-3	12,082.7	3,852.4	66,465.7	728,130.0
	1E-1	12,496.3	3,894.6	68,283.8	733,089.2
64	1E-5	12,440.5	3,900.4	68,675.4	733,853.3
	1E-3	12,443.4	3,900.6	68,685.1	733,954.1
	1E-1	12,790.1	3,924.3	69,633.8	737,897.4
128	1E-5	13,128.1	3,964.0	71,020.3	742,402.1
	1E-3	13,130.1	3,964.0	71,018.8	742,431.6
	1E-1	13,359.4	3,977.0	71,515.9	745,671.6

eTable 1: Mean squared error between reported data and incidence curve convolved with delay distribution by back projection parameter value.

Virginia. Hospitalization data are from the Virginia Department of Health https://www.vdh.virginia.gov/coronavirus/. Regions are state health regions as defined by that department.

Delay distributions.

We model delay distributions as a sum of delay times. First, we fit a distribution from infection to onset of symptoms. Next, we fit three separate distributions for: onset of symptoms to test report date; onset of symptoms to hospitalization date; and hospitalization date to death. The final delay distribution is found by composing the infection to symptom onset distribution and the appropriate symptom onset to recorded event distribution. We model all distributions with gamma distributions to capture observed unimodality and long right tails with only two parameters.

Uncertainty in delay distributions comes from a number of areas. First, many of these data sets are relatively small; the hospital admittance to death distribution is based on 202 records [5]. Second, most published data only captures cases from early to mid-2020. It is conceivable that time until death may have changed based on treatment advances. We did examine changes in

State	Region	Start Date	End Date	Cases	Hospitalizations	Deaths
NY	State	2020-03-03	2020-12-18	832,032	—	36,945
	Brooklyn	2020-03-03	2020-12-18	99,786	18,462	5,980
	Bronx	2020-03-03	2020-12-18	71,227	14,014	4,177
	Manhattan	2020-03-03	2020-12-18	48,386	9,164	2,635
	Queens	2020-03-03	2020-12-18	101,370	19,821	6,245
	Staten Island	2020-03-03	2020-12-18	27,299	3,309	1,042
	NYC	2020-03-03	2020-12-18	348,068	64,770	20,079
AZ	State	2020-01-26	2020-11-30	327,048	27,581	6,640
OH	State	2020-03-09	2020-12-05	467,433	27,954	6,947
TX	State	2020-02-12	2020-12-21	1,610,141	—	26,165
	Houston	2020-03-04	2020-12-21	317,327	—	4,313
	Dallas/Fort Worth	2020-03-09	2020-12-21	434,283	—	4,560
	Austin	2020-03-13	2020-12-21	81,703	—	917
VA	State	2020-03-17	2020-12-22	314,481	17,083	4,705
	Central	2020-03-17	2020-12-22	51,261	2,848	923
	Eastern	2020-03-17	2020-12-22	60,668	3,503	846
	Northern	2020-03-17	2020-12-22	95,367	5,916	1,367
	Northwest	2020-03-17	2020-12-22	47,431	2,180	577
	Southwest	2020-03-17	2020-12-22	59,754	2,636	992

eTable 2: Summary of data date ranges and summary statistics by region and state.

time from symptoms until a recorded test using line data from the Florida Department of Health from April 2020 through July 14, 2020 [15]; all monthly variations from the mean were on the order of about one day. See eFigure 11 for a comparison of incidence fits using subsets of line data where testing delay distribution contributions are fit by month.

Infection to symptom onset. We estimate time from infection to symptom onset by matching quantiles of a gamma distribution to the data from [16]; this had a shape = 3.57 and a rate = 0.7. The mean incubation period was 5.1 days and 97.5% had shown symptoms by 11.5 days.

Symptom onset to recorded test. Symptom onset to positive test report by fitting a gamma distribution to non-zero delay times from Florida for all cases through July 14, 2020 [15]. Data was cleaned by removing all zero records, which was 50.5% of the data, and capping all recording delays at 40 days from symptom onset. This led to a mean of 6.65 days between symptom onset and recorded test date, which was fit with a gamma distribution with parameters shape = 1.3101 and rate = 0.1969.

Symptom onset to hospitalization. The hospitalization delay distribution is the symptom delay distribution plus a delay distribution between symptoms and hospitalization, which [5] modeled as a gamma distribution with shape = 5.078 and rate = 1.307. This uses data from [17, 18].

Hospitalization to death. We fit a gamma distribution to the time from hospital admission until death for the distributions presented in [5] using quantile matching. This had a mean of 14.5 days, a median of 12.7 days, and a 25%–75% centile range of 7.3 to 19.9 days; giving shape

= 2.14 and rate = 0.15. We note that this data only represents deaths that occurred after hospital admission. We do not believe this to be a limiting factor for data quality as high proportions of COVID-19 diagnosed deaths occurred within the hospital system; 95.5% of COVID-19 diagnosed deaths occurred in hospital in the Mass General Brigham health system between February 18 and May 18, 2020 [19].

Estimating infections at state and local scales

COVID-19 data — cases, deaths, and hospitalizations — are often reported at the state or national level, which obscures local variation in prevalence and response to policy interventions. We compare inferred infections from cases in the five New York City boroughs with New York state; the Austin, Dallas/Fort Worth, and Houston metros with the state of Texas; and Virginia health regions (Central, Eastern, Northern, Northwest, and Southwest) to statewide across Virginia. Inferred infections are shown in eFigure 12. For each region and state, we also include the following policy decisions: school closures, gathering bans for 10 people or more, stay-at-home orders, reopening of low risk non-essential businesses, and reopening of bars or restaurants with indoor seating. Since these restrictions were added and lifted early in the pandemic, we estimated incidence through June 30, 2020.

Given the length of the time series fitted and the noise of the case data, slope changes in infections are accurate to within a few days due to estimator smoothing. Our estimates suggest an association between indoor bar/restaurant openings and an increase in transmission in Texas and the southwestern and eastern regions of Virginia. Additionally, we find that in New York City, estimates suggest that the infections incidence decline postdated school closures, but was aligned with statewide stay-at-home orders. Likewise, statewide — but not county level — stay-at-home orders coincided with a local incidence decline in Houston in early April. Earlier stay-at-home orders in Harris county were not aligned with the decline (eFigure 12).

References

[1] Open COVID-19 Data Working Group. Detailed Epidemiological Data from the COVID-19 Outbreak. Accessed on 2020-06-16 from http://virological.org, 2020.

[2] Bo Xu, Bernardo Gutierrez, Sumiko Mekaru, Kara Sewalk, Lauren Goodwin, Alyssa Loskill, Emily Cohn, Yulin Hswen, Sarah C. Hill, Maria M Cobo, Alexander Zarebski, Sabrina Li, Chieh-Hsi Wu, Erin Hulland, Julia Morgan, Lin Wang, Katelynn O'Brien, Samuel V. Scarpino, John S. Brownstein, Oliver G. Pybus, David M. Pigott, and Moritz U. G. Kraemer. Epidemiological data from the COVID-19 outbreak, real-time case information. *Scientific Data*, 7(106), 2020. doi: doi.org/10.1038/s41597-020-0448-0.

[3] Sam Abbott, Joel Hellewell, Robin N Thompson, Katharine Sherratt, Hamish P Gibbs, Nikos I Bosse, James D Munday, Sophie Meakin, Emma L Doughty, June Young Chun, et al. Estimating the time-varying reproduction number of sars-cov-2 using national and subnational case counts. *Wellcome Open Research*, 5(112):112, 2020.

[4] Sam Abbott, Joel Hellewell, Robin N. Thompson, Katherine Sherratt, Hamish P Gibbs, Nikos I Bosse, James D Munday, Sophi Meakin, Emma L Doughty, June Young Chun, Yung-Wai Desmond Chan, Flavio Finger, Paul Campbell, Aira Endo, Carl A B Pearson, Amy Gimma, Tim Russel, CMMID COVID modeling group, Stefan Flasche, Adam J Kucharski, Rosalind M Eggo, and Sebastian Funk. Temporal variation in transmission during the COVID-19 outbreak. https://epiforecasts.io/covid/ (accessed 2020-09-01), 2020. URL https://epiforecasts.io/covid/.

[5] Joseph A Lewnard, Vincent X Liu, Michael L Jackson, Mark A Schmidt, Britta L Jewell, Jean P Flores, Chris Jentz, Graham R Northrup, Ayesha Mahmud, Arthur L Reingold, et al. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: prospective cohort study. *BMJ*, 369, 2020.

[6] Timothy W Russell, Joel Hellewell, Christopher I Jarvis, Kevin Van Zandvoort, Sam Abbott, Ruwan Ratnayake, Stefan Flasche, Rosalind M Eggo, W John Edmunds, Adam J Kucharski, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance*, 25(12):2000256, 2020.

[7] Kevin Systrom and Thomas Vladeck. R_t Covid-19. https://rt.live (accessed 2020-06-16), 2020.

[8] Finbarr O'Sullivan. A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1(4):502–518, 1986.

[9] Ron Brookmeyer and Mitchell H Gail. A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *Journal of the American Statistical Association*, 83(402):301–308, 1988.

[10] Niels G Becker, Lyndsey F Watson, and John B Carlin. A method of non-parametric back-projection and its application to AIDS data. *Statistics in Medicine*, 10(10):1527–1542, 1991.

[11] Bradley Efron. Empirical Bayes deconvolution estimates. *Biometrika*, 103(1):1–20, 2016.

[12] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

[13] Francis J Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254, 1948.

[14] Leonardo S Bastos, Theodoros Economou, Marcelo FC Gomes, Daniel AM Villela, Flavio C Coelho, Oswaldo G Cruz, Oliver Stoner, Trevor Bailey, and Claudia T Codeço. A modelling approach for correcting reporting delays in disease surveillance data. *Statistics in Medicine*, 38(22):4363–4377, 2019.

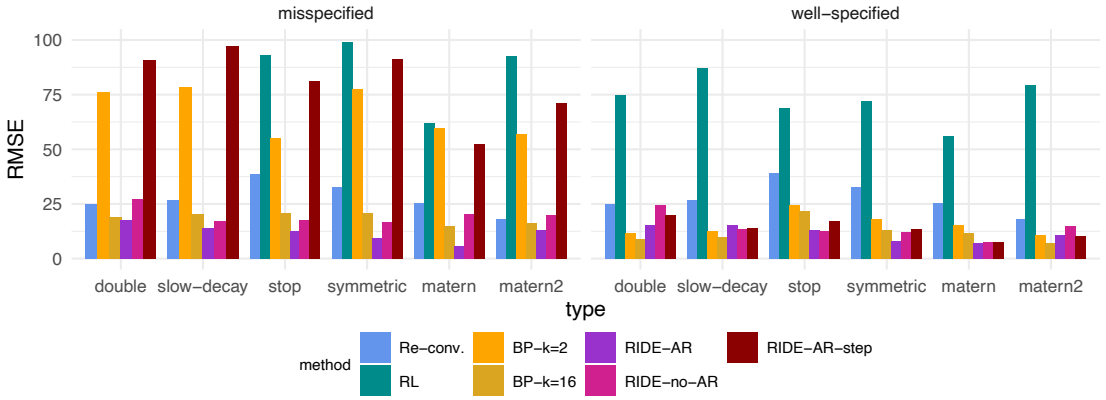
[15] Florida Department of Health. Florida COVID19 case line data. https://open-fdoh.hub.arcgis.com/datasets/florida-covid19-case-line-data, 2020. URL https://open-fdoh.hub.arcgis.com/datasets/florida-covid19-case-line-data. Accessed: 2020-07-15.

[16] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, 172(9):577–582, 2020.

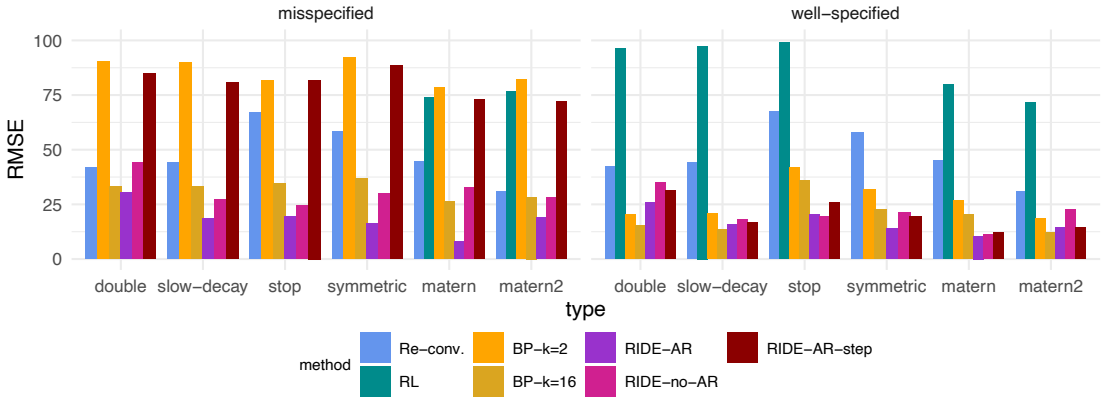
[17] Dawei Wang, Bo Hu, Chang Hu, Fangfang Zhu, Xing Liu, Jing Zhang, Binbin Wang, Hui Xiang, Zhenshun Cheng, Yong Xiong, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *Journal of the American Medical Association*, 323(11):1061–1069, 2020.

[18] Juanjuan Zhang, Maria Litvinova, Wei Wang, Yan Wang, Xiaowei Deng, Xinghui Chen, Mei Li, Wen Zheng, Lan Yi, Xinhua Chen, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases*, 2020.

[19] Isaac S Chua, Sandra M Shi, and David M Levine. Place of death and end-of-life care utilization among covid-19 decedents in a massachusetts health care system. *Journal of Palliative Medicine*, 24(3):322–323, 2021.

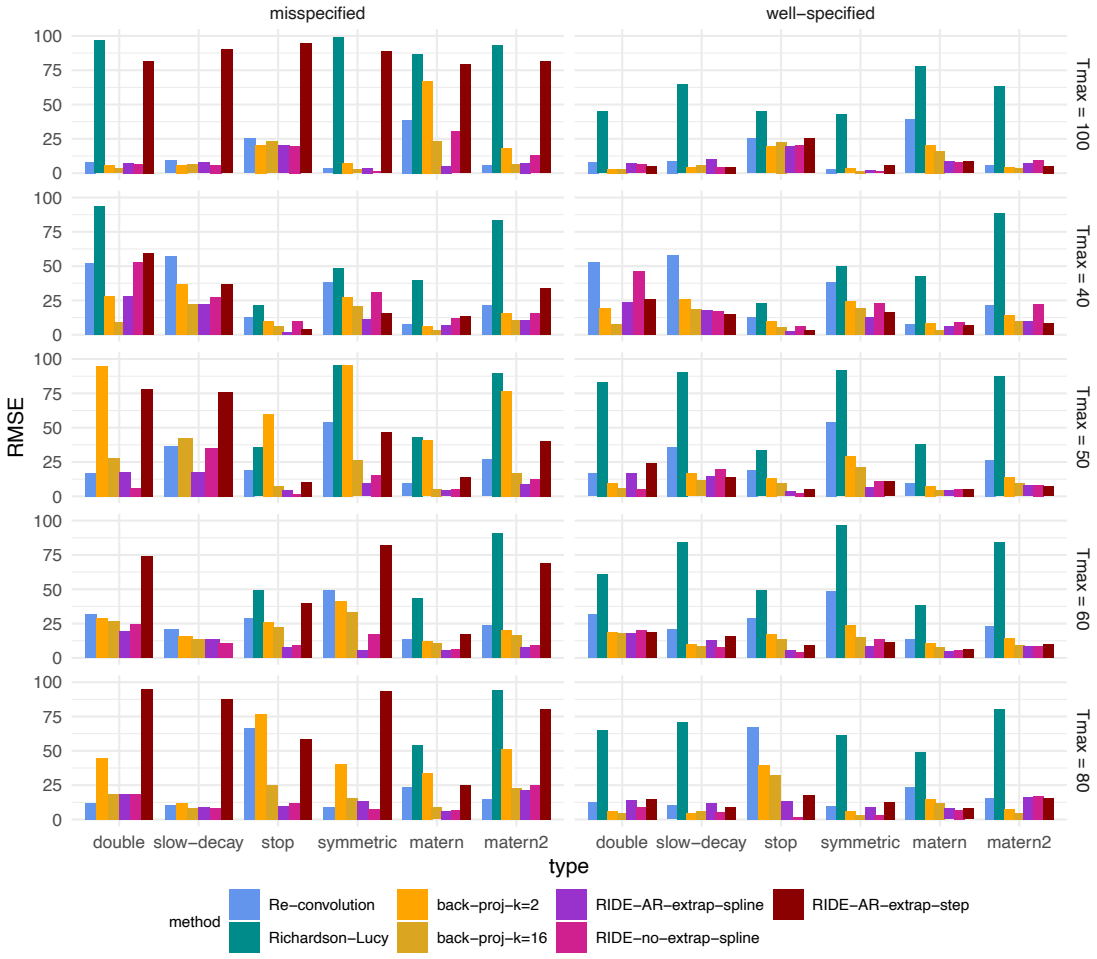


(a) Average RMSE over all experimental settings.

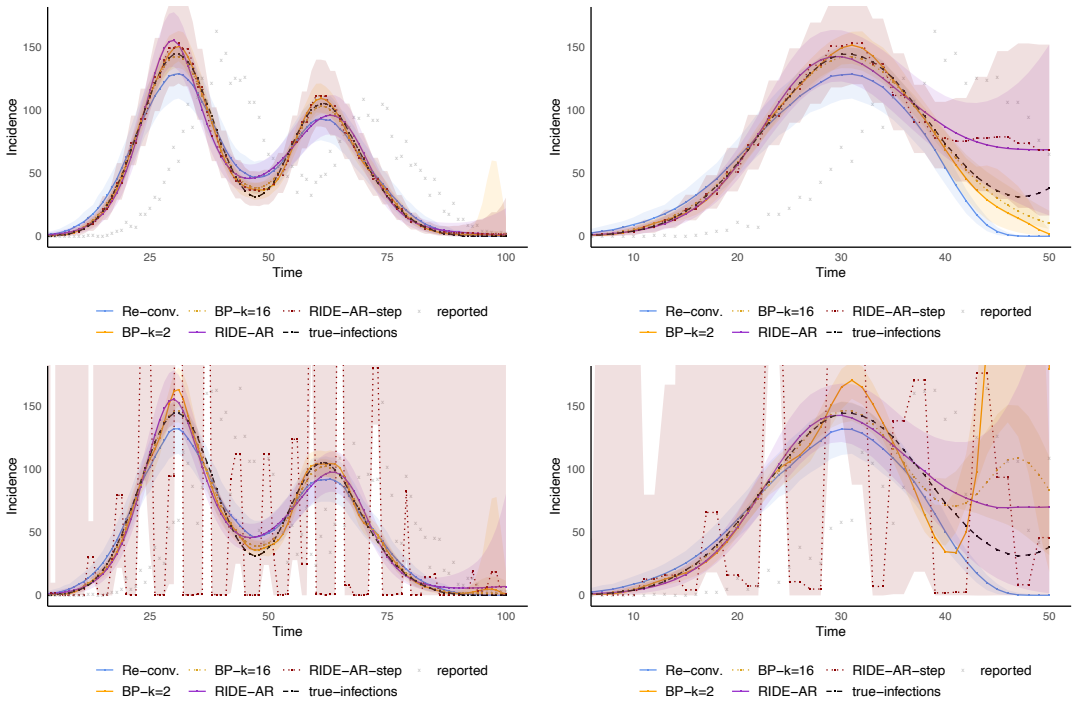


(b) Average RMSE for the most recent 20 days, averaged over all experimental settings.

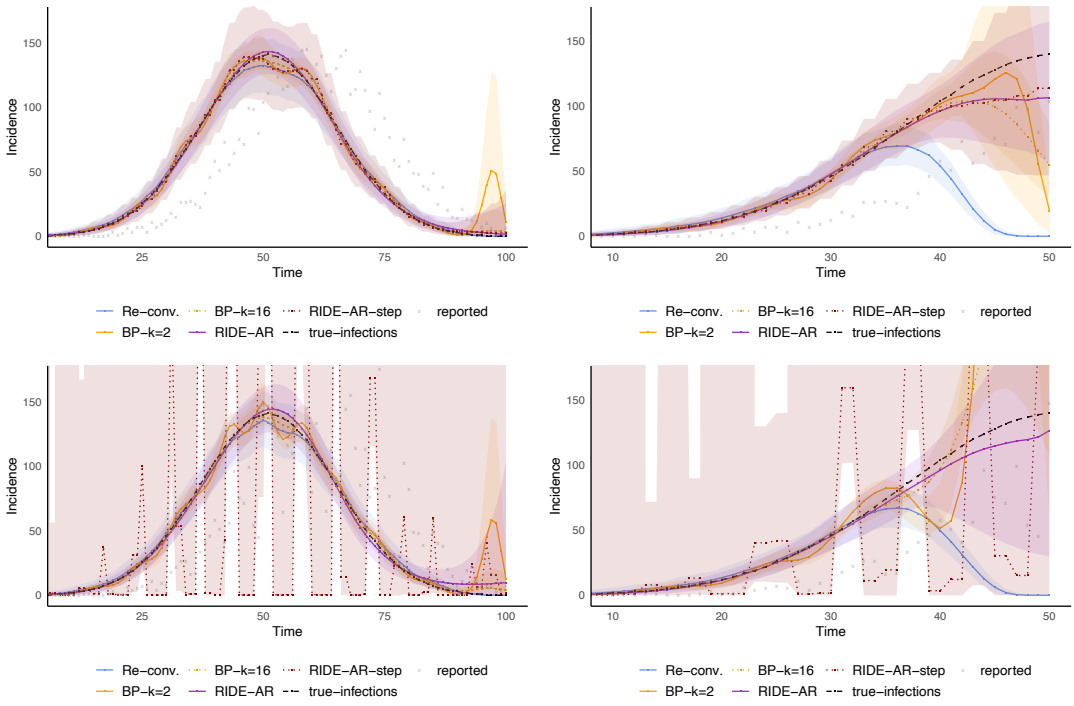
eFigure 1: Synthetic experiment error — comparison of RMSE by method, curve type, and noise type (misspecified and well-specified delay models), comparing (a) overall average error, and (b) average error on most recently observed 20 days. Methods with average error greater than 100 are not depicted.



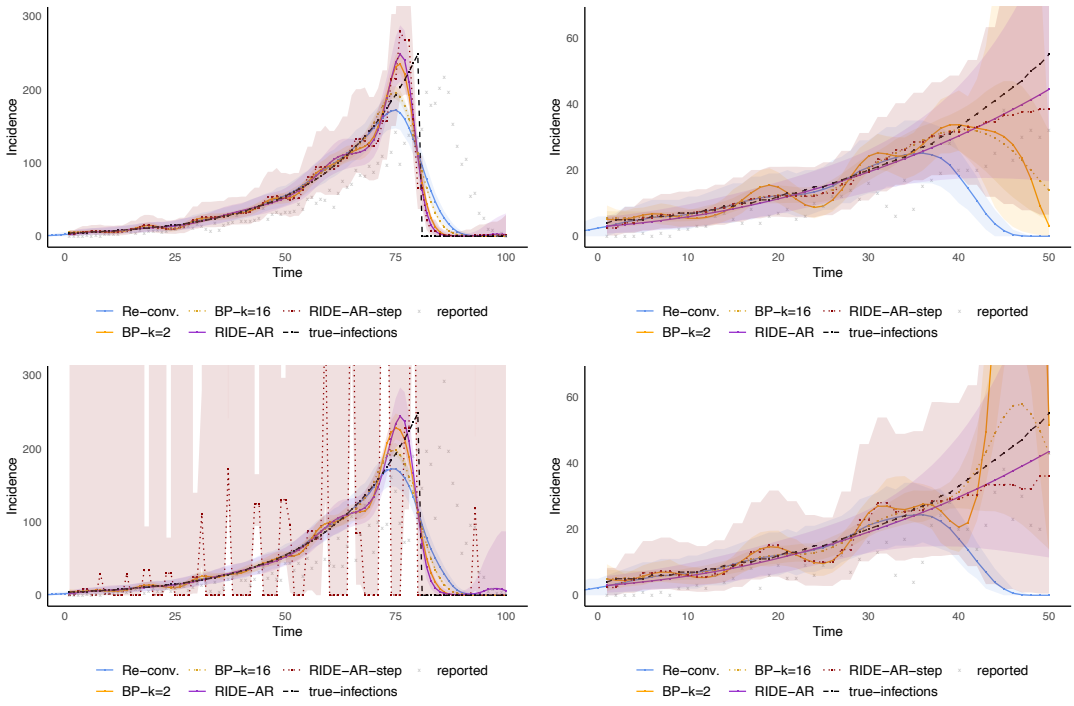
eFigure 2: Synthetic experiment error — comparison of RMSE by method, varying censor time (T_{max}) and noise type (misspecified and well-specified delay models).



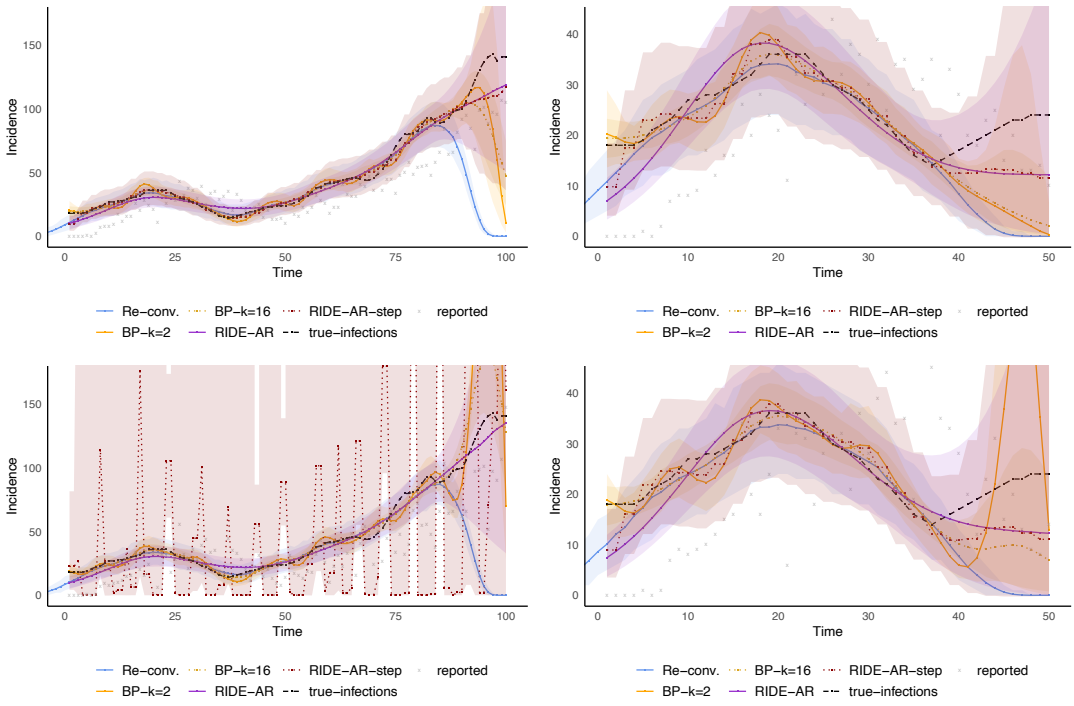
eFigure 3: `double`: Top: correctly specified full (left) and censored (right). Bottom: misspecified full (left) and censored (right). We observe that the spline model copes with misspecified observations more robustly.



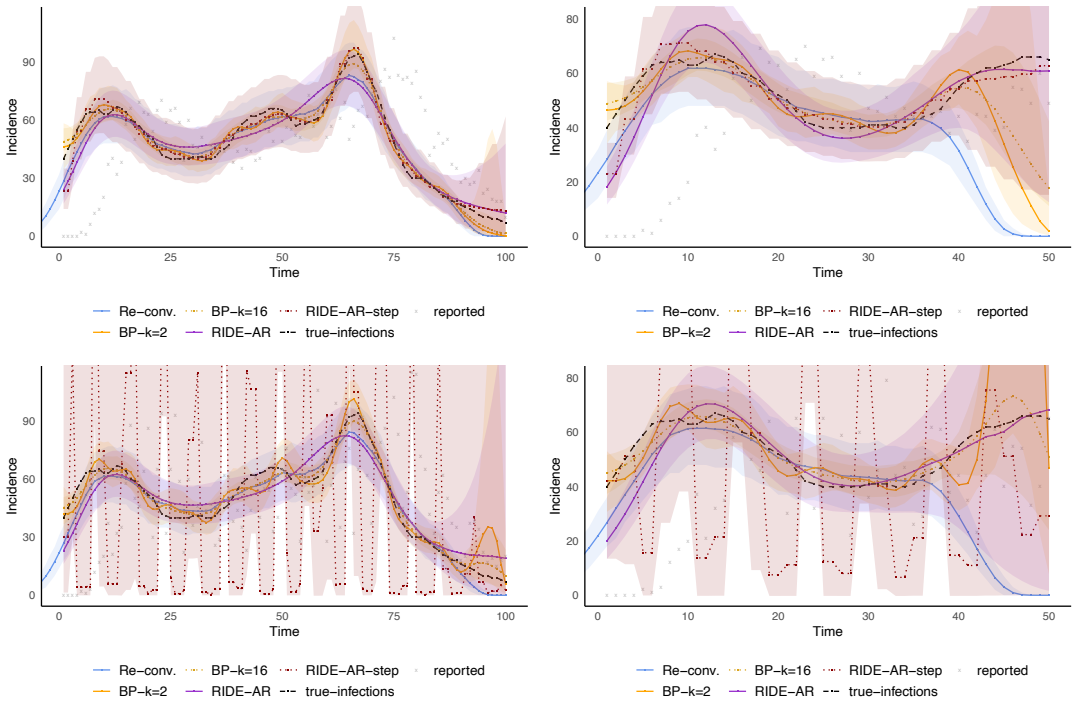
eFigure 4: `symmetric`: Top: correctly specified full (left) and censored (right). Bottom: misspecified full (left) and censored (right). Again, we observe that the spline model copes with misspecified observations more robustly.



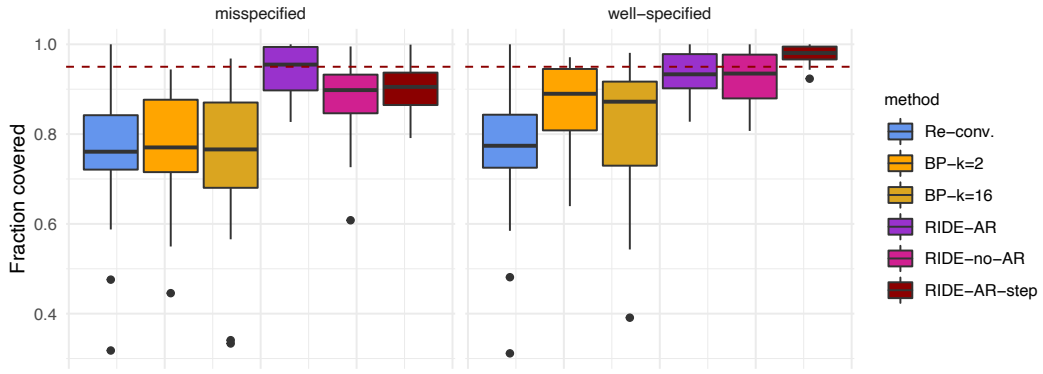
eFigure 5: `stop`: Top: correctly specified full (left) and censored (right). Bottom: misspecified full (left) and censored (right). Again, we observe that the spline model copes with misspecified observations more robustly.



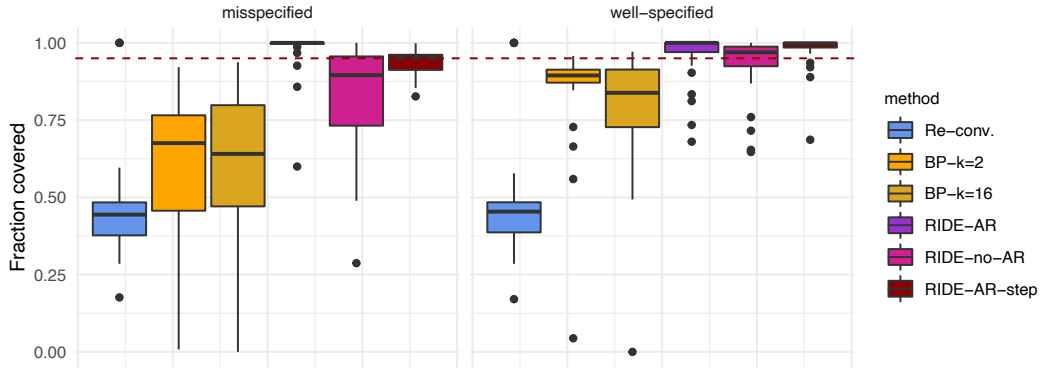
eFigure 6: `matern`: Top: correctly specified full (left) and censored (right). Bottom: misspecified full (left) and censored (right). Again, we observe that the spline model copes with misspecified observations more robustly.



eFigure 7: `matern2`: Top: correctly specified full (left) and censored (right). Bottom: misspecified full (left) and censored (right). Again, we observe that the spline model copes with misspecified observations more robustly.

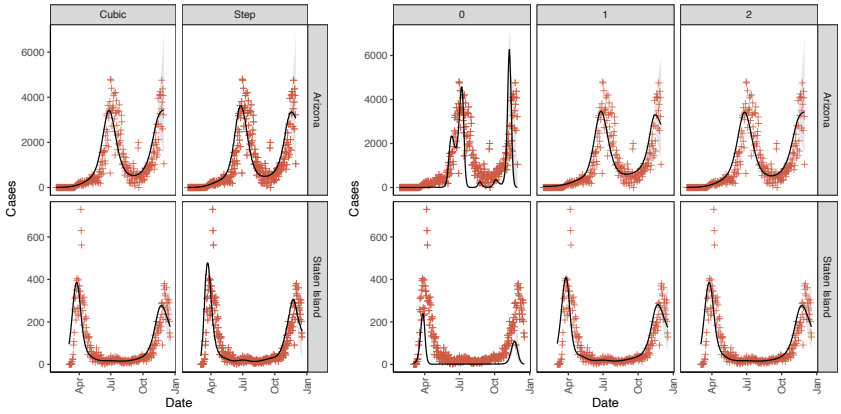


(a) Total average coverage.

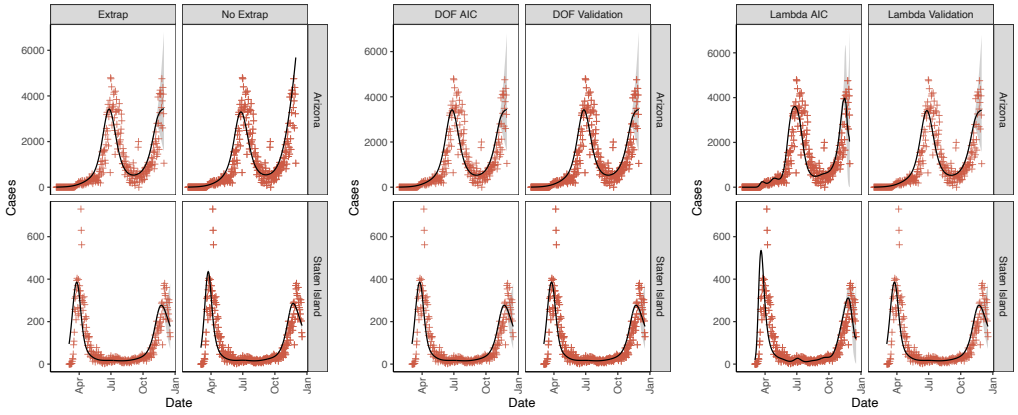


(b) Average coverage on recent observations (last 20 days).

eFigure 8: Average coverage at the 95% level (dashed red line) across all experimental settings, replicated 50 times. Re-convolution and back projection (for the default $k = 2$ and smoothed $k = 16$) underestimate uncertainty marginally. Note that in more recent observations (i.e., in the previous 20 days), re-convolution and back projection coverage becomes worse on average, while the RIDE coveage remains relatively fixed (if not a little conservative).

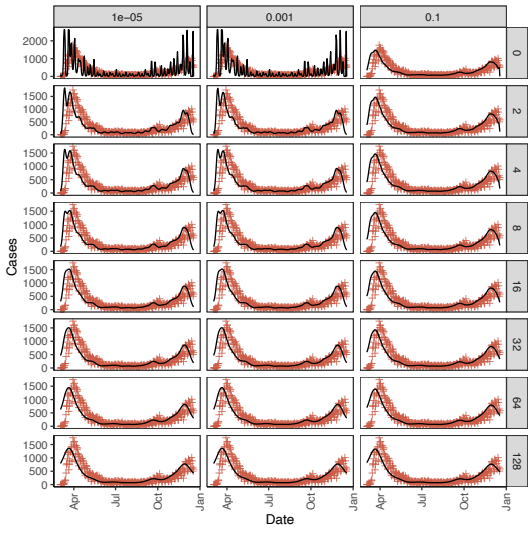


(a) Basis comparison. (b) Regularization order.

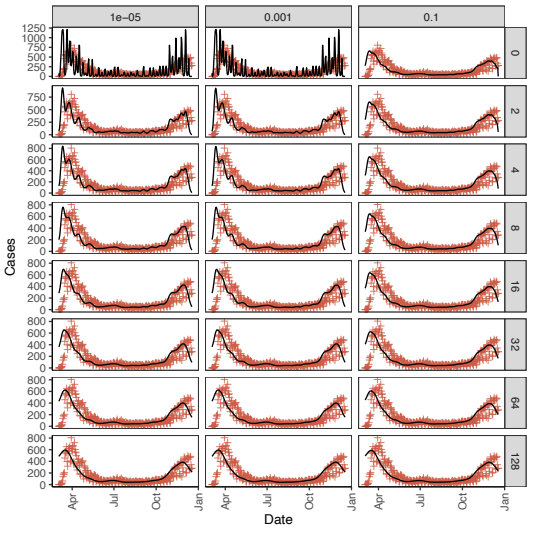


(c) Extrapolation. (d) DOF selection. (e) Lambda selection.

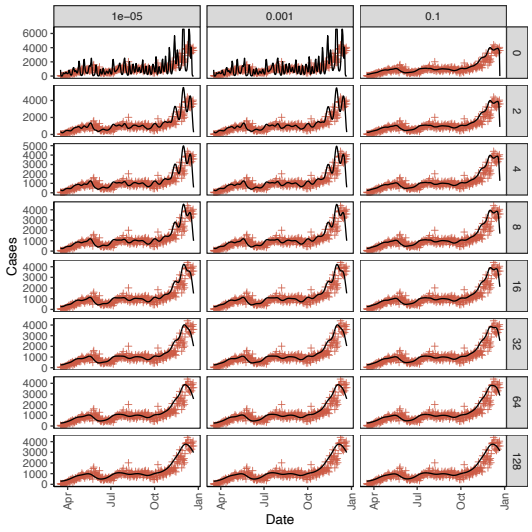
eFigure 9: Estimated infections incidence (solid black line), 90% credible regions (gray shaded) when available, and observed values (red plus) by data type across regions (rows) and by parameter value (columns). (a) Basis comparison: cubic splines (default) vs. daily step functions. (b) Regularization order comparison: 0, 1, and 2 (default). (c) Extrapolation comparison: extrapolation (default) vs. no extrapolation. (d) Cubic spline degree of freedom (dof) selection method: AIC (default) vs. validation. (e) Regularization parameter (λ) selection method: validation (default) vs. AIC.



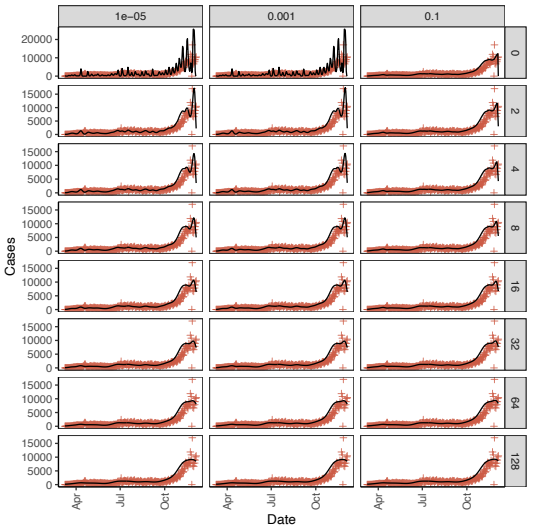
(a) Brooklyn cases.



(b) Manhattan cases.

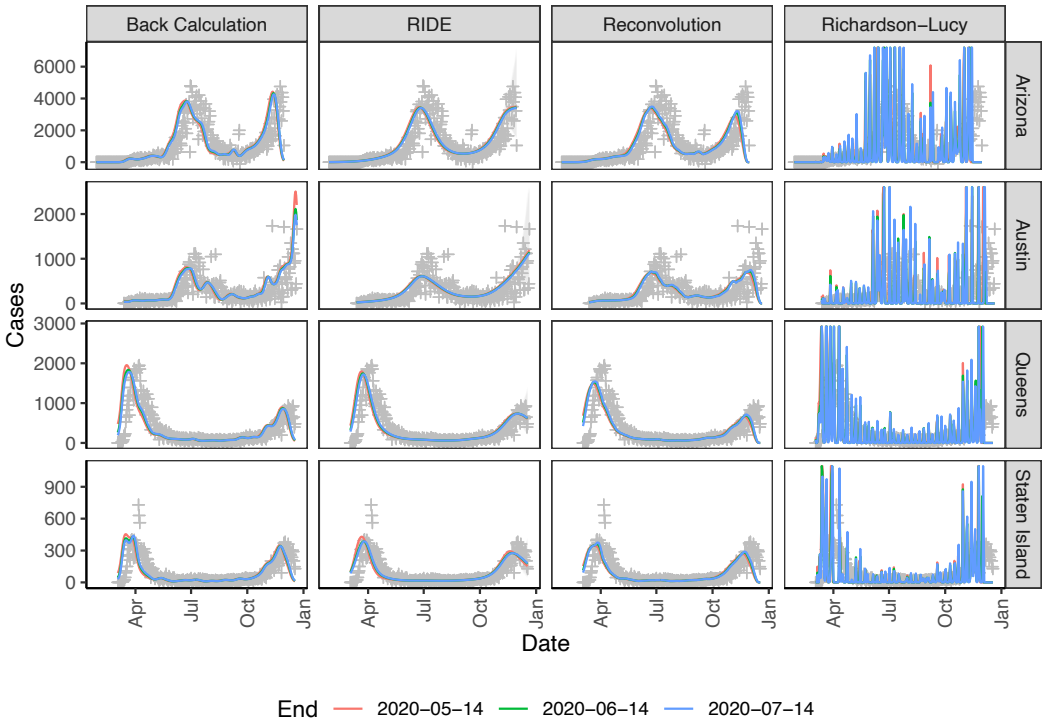


(c) Virginia cases.

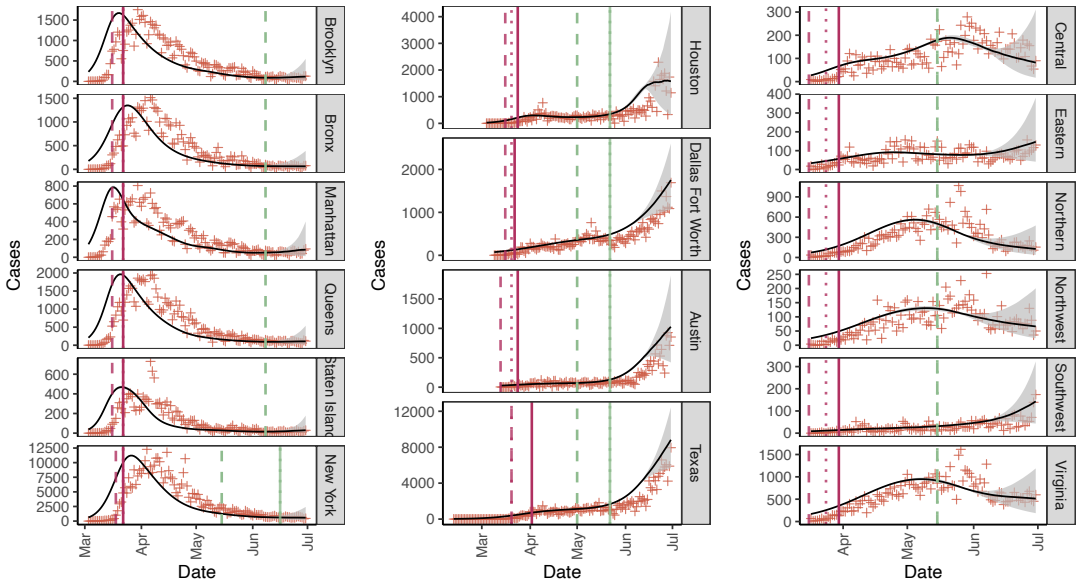


(d) Ohio cases.

eFigure 10: Comparison of back projection fits across values of k (rows) and epsilon (columns), for (a) Brooklyn, (b) Manhattan, (c) the state of Virginia, and (d) the state of Ohio.



eFigure 11: Incidence fits from recorded cases across method using line data [15] subsets to calculate time from symptoms to recorded test: March 1, 2020 through May 14, 2020; May 15, 2020 through June 14, 2020; and June 15, 2020 through July, 14. Although variation in testing capacity and speed has changed throughout the pandemic, differences in fits are minimal. All other parts of the delay distribution are generated from summary statistics rather than raw data.



eFigure 12: Incidence (solid black line), 90% credible regions (gray shaded), and observed values (red plus) fitted on case data for New York, Texas, and Virginia by sub-region. Policy dates are included: school closures (red thick dash), gatherings ban for 10 or more people (red thin dash), stay-at-home order (red solid line), reopening of low risk non-essential businesses (green thick dash), and bars/restaurants with indoor seating (green solid line).