

RtEstim: Effective reproduction number estimation with trend filtering

Jiapeng Liu^{1*}, Zhenglun Cai², Paul Gustafson¹, Daniel J. McDonald¹

1 Department of Statistics, The University of British Columbia, Vancouver, British Columbia, Canada

2 Centre for Health Evaluation and Outcome Sciences, The University of British Columbia, Vancouver, British Columbia, Canada

* jiaping.liu@stat.ubc.ca

Abstract

To understand the transmissibility and spread of infectious diseases, epidemiologists turn to estimates of the effective reproduction number. While many estimation approaches exist, their utility may be limited. Challenges of surveillance data collection, model assumptions that are unverifiable with data alone, and computationally inefficient frameworks are critical limitations for many existing approaches. We propose a discrete spline-based approach **RtEstim** that solves a convex optimization problem—Poisson trend filtering—using the proximal Newton method. It produces a locally adaptive estimator for effective reproduction number estimation with heterogeneous smoothness. **RtEstim** remains accurate even under some process misspecifications and is computationally efficient, even for large-scale data. The implementation is easily accessible in a lightweight R package **rtestim**.

Author summary

Effective reproduction number estimation presents many challenges due to data collection, modelling assumptions, and computational burden. Such limitations hinder the accurate estimation of the effective reproduction number. Our motivation is to

develop a model that produces accurate estimates, is robust to model misspecification, and is straightforward to use and computationally efficient, even for large counts and long time periods. We propose a convex optimization model with an ℓ_1 trend filtering penalty. It couples accurate estimation of the effective reproduction number with desired smoothness. We solve the optimization using the proximal Newton method, which converges rapidly and is numerically stable. Our software, conveniently available in the R package `RtEstim`, can produce estimates in seconds for incidence sequences with hundreds of observations. These estimates are produced for a sequence of tuning parameters and can be selected using a built-in cross validation procedure.

1 Introduction

The effective reproduction number at time t is defined to be the expected number of secondary infections produced by a primary infection throughout the course of the entire infection if conditions remain the same at the specific time. The instantaneous reproduction number, specifically, is a type of effective reproduction number focusing on the transmission at a specific timepoint [1]. It is a key quantity for understanding infectious disease dynamics including the potential size of an outbreak and the required stringency of control measures [2,3]. Tracking the time series of this quantity is useful for understanding whether or not future infections are likely to increase or decrease from the current state [4]. Let $\mathcal{R}(t)$ denote the effective reproduction number at time t . Practically, as long as $\mathcal{R}(t) < 1$, infections will decline gradually, eventually resulting in a disease-free equilibrium, whereas when $\mathcal{R}(t) > 1$, infections will continue to increase, resulting in endemic equilibrium. While $\mathcal{R}(t)$ is fundamentally a continuous time quantity, it can be related to data only at discrete points in time $t = 1, \dots, n$. This sequence of effective reproduction numbers over time is not observable, but, nonetheless, is easily interpretable and retrospectively describes the course of an epidemic. Therefore, a number of procedures exist to estimate \mathcal{R}_t from different types of observed incidence data such as cases, deaths, or hospitalizations, while relying on various domain-specific assumptions, e.g., [5–8]. Importantly, accurate estimation of effective reproduction numbers relies heavily on the quality of the available data, and, due to the limitations of data collection, such as underreporting and lack of standardization, estimation

methodologies rely on various assumptions to compensate. Because model assumptions
may not be easily verifiable from data alone, it is also critical for any estimation
procedure to be robust to model misspecification.

Many existing approaches for effective reproduction number estimation are Bayesian:
they estimate the posterior distribution of \mathcal{R}_t conditional on the observations. One of
the first such approaches is the software **EpiEstim** [9], described in [10]. This method is
prospective, in that it uses only observations available up to time t in order to estimate
 \mathcal{R}_t for each $i = 1, \dots, t$. An advantage of **EpiEstim** is its straightforward statistical
model: new incidence data follows the Poisson distribution conditional on past incidence
combined with the conjugate gamma prior distribution for \mathcal{R}_t with fixed
hyperparameters. Additionally, the serial interval distribution, the distribution of the
period between onsets of primary and secondary infections in a population, is fixed and
known. For this reason, **EpiEstim** requires little domain expertise for use, and it is
computationally fast. [11] modified this method to distinguish imported cases from local
transmission and simultaneously estimate the serial interval distribution. [12] further
extended **EpiEstim** by using “reconstructed” daily incidence data to handle irregularly
spaced observations. Recently, [13] proposed a Bayesian latent variable framework,
EpiNow2 [14], which leverages incident cases, deaths or other available streams
simultaneously along with allowing additional delay distributions (incubation period
and onset to reporting delays) in modelling. [15] proposed an extension that handles
missing data by imputation followed by a truncation adjustment. These modifications
are intended to increase accuracy at the most recent (but most uncertain) timepoints,
to aid policymakers. [16] also proposed a Bayesian approach, **EpiFilter**, based on the
(discretized) Kalman filter and smoother. **EpiFilter** also estimates the posterior of \mathcal{R}_t
given a Gamma prior and Poisson distributed incident cases. Compared to **EpiEstim**,
however, **EpiFilter** estimates \mathcal{R}_t retrospectively using all available incidence data both
before and after time t , with the goal of being more robust in low-incidence periods. [17]
proposed a Bayesian P-splines approach, **EpiLPS**, that assumes negative Binomial
distributed observations. [18] also proposed a Bayesian model estimated with particle
filtering to incorporate spatial structures. Bayesian approaches estimate the posterior
distribution of the effective reproduction numbers and possess the advantage that
credible intervals may be easily computed. They incorporate the prior knowledge on

parameters to modelling. Some techniques are used to eliminate the power of prior
54 parameters on the posterior estimates to make the estimates more plausible, e.g., [11]
55 assumes a relatively large prior mean of \mathcal{R}_t (appreciably larger than 1), and if the
56 estimate is less than 1, researchers will know it is a direct result from data, instead of
57 the choice of prior parameters. Some Bayesian approaches, however, are
58 computationally expensive, since they require more intensive computational routines,
59 especially when observed data sequences are long or hierarchical structures are complex,
60 e.g., [13]. While, some Bayesian methods with more efficient structures, especially the
61 ones with only conjugate priors, can be computationally efficient, e.g., [10]. Below, we
62 compare our method to the aforementioned Bayesian models, **EpiEstim**, **EpiLPS**,
63 **EpiFilter**, and **EpiNow2**.
64

There are also frequentist approaches for \mathcal{R}_t estimation. [19] proposed regularizing
65 the smoothness of \mathcal{R}_t through penalized regression with second-order temporal
66 regularization, additional spatial penalties, and with Poisson loss. [20] extended this
67 procedure by introducing another penalty on outliers. [21] proposed a spline-based
68 model relying on the assumption of exponential-family distributed incidence. [22]
69 estimates \mathcal{R}_t while monitoring the time-varying level of overdispersion. There are other
70 spline-based approaches such as [23, 24], autoregressive models with random effects [25]
71 that are robust to low incidence, and generalized autoregressive moving average
72 (GARMA) models [26] that are robust to measurement errors in incidence data.
73

We propose an estimator for retrospective estimation of effective reproduction
74 number, instantaneous reproduction number specifically, called **RtEstim** that requires
75 only incidence data. Our model makes the conditional Poisson assumption, similar to
76 much of the prior work described above, but is empirically more robust to
77 misspecification. This estimator is defined by a convex optimization problem with
78 Poisson loss and ℓ_1 penalty on the temporal evolution of $\log(\mathcal{R}_t)$ to impose smoothness
79 over time. As a result, **RtEstim** generates discrete splines, and the estimated curves (in
80 logarithmic space) appear to be piecewise polynomials of an order selected by the user.
81 Importantly, the estimates are locally adaptive, meaning that different time ranges may
82 possess heterogeneous smoothness. Because we penalize the logarithm of \mathcal{R}_t , we
83 naturally accommodate the positivity requirement, in contrast to related methods, can
84 handle large or small incidence measurements, and are automatically (reasonably)
85

robust to outliers without additional constraints. A small illustration using three years of Covid-19 case data in Canada is shown in Fig 1 [27].

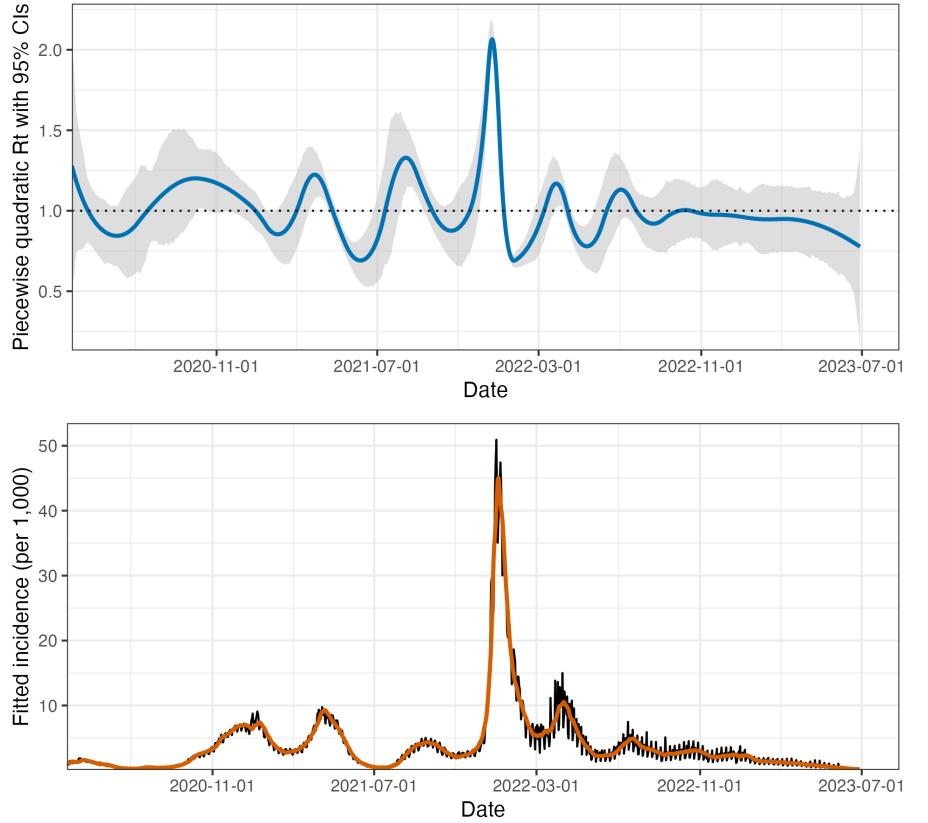


Fig 1. A demonstration of effective reproduction number estimation by RtEstim and the corresponding predicted incident cases for the Covid-19 epidemic in Canada during the period from March 28, 2020 to June 28, 2023. In the top panel, the blue curve is the estimated piecewise quadratic R_t and the gray ribbon is the corresponding 95% confidence band. The black curve in the bottom panel is the observed Covid-19 daily confirmed cases, and the orange curve is the predicted incident cases corresponding to the estimated R_t .

While our approach is straightforward and requires little domain knowledge for implementation, we also implement a number of refinements. We use a proximal Newton method to solve the convex optimization problem along with warm starts to produce estimates efficiently, typically in a matter of seconds, even for long sequences of data. In a number of simulation experiments, we show empirically that our approach is more accurate than existing methods at estimating the true effective reproduction numbers and robust under multiple settings of the misspecification of incidence distribution, serial interval distribution, and the order of graphical curvature.

The manuscript proceeds as follows. We first introduce the methodology of RtEstim

including the renewal equation and the development of Poisson trend filtering estimator. 97
 We explain how this method could be interpreted from the Bayesian perspective, 98
 connecting it to previous work in this context. We provide illustrative experiments 99
 comparing our estimator to other Bayesian alternatives. We then apply our RtEstim 100
 and the alternatives on the Covid-19 pandemic incidence in British Columbia and the 101
 1918 influenza pandemic incidence in the United States. Finally, we conclude with a 102
 discussion of the advantages and limitations of our approach and describe practical 103
 considerations for effective reproduction number estimation. 104

2 Methods

2.1 Renewal model for incidence data

The effective reproduction number $\mathcal{R}(t)$ is defined to be the expected number of 107
 secondary infections at time t produced by a primary infection sometime in the past. To 108
 make this precise, denote the number of new infections at time t as $y(t)$. Then the total 109
 primary infectiousness can be written as $\eta(t) := \int_0^\infty p(i)y(t-i)di$, where $p(i)$ is the 110
 probability that a new secondary infection is the result of a primary infection that 111
 occurred i time units in the past. The effective reproduction number is then given as 112
 the value that equates 113

$$\mathbb{E}[y(t) | y(j), j < t] = \mathcal{R}(t)\eta(t) = \mathcal{R}(t) \int_0^\infty p(i)y(t-i)di, \quad (1)$$

otherwise known as the renewal equation. The period between primary and secondary 114
 infections is exactly the generation time of the disease, but given real data, observed at 115
 discrete times (say, daily), this delay distribution must be discretized into contiguous 116
 time intervals, say, $(0, 1], (1, 2], \dots$. It results in the sequence $\{p_i\}_0^\infty$ corresponding to 117
 observations y_t and yields the discretized version of Eq (1), 118

$$\mathbb{E}[y_t | y_1, \dots, y_{t-1}] = \mathcal{R}_t \eta_t = \mathcal{R}_t \sum_{i=1}^{\infty} p_i y_{t-i}. \quad (2)$$

Many approaches to estimating \mathcal{R}_t rely on Eq (2) as motivation for their procedures, 119
 among them, EpiEstim [10] and EpiFilter [16]. 120

In most cases, it is safe to assume that infectiousness disappears beyond τ
timepoints ($p(i) = 0$ for $i > \tau$), resulting in the truncated integral of the generation
interval distribution $\int_0^\tau p(i)di = 1$. Generation time, however, is usually unobservable
and tricky to estimate, so common practice is to approximate it by the serial interval:
the period between the symptom onsets of primary and secondary infections. If the
infectiousness profile after symptom onset is independent of the incubation period (the
period from the time of infection to the time of symptom onset), then this
approximation is justifiable: the serial interval distribution and the generation interval
distribution share the same mean. However, other properties may not be similarly
shared, and, in general, the generation interval distribution is a convolution of the serial
interval distribution with the distribution of the difference between independent draws
from the delay distribution from infection to symptom onset. See, for example, [1] for a
fuller discussion of the dangers of this approximation. Nonetheless, treating these as
interchangeable is common [10] and doing otherwise is beyond the scope of this work.
Additionally, we assume that the generation interval (and, therefore, the serial interval),
is constant over time t . That is, the probability $p(i)$ depends only on the gap between
primary and secondary infections and not on the time t when the secondary infection
occurs. For our methods, we will assume that the serial interval can be accurately
estimated from auxiliary data (say by contact tracing, or previous epidemics) and we
will take it as fixed, as is common in existing studies, e.g., [10, 19, 20].

The renewal equation in Eq (2) relates observable data streams (incident cases)
occurring at different timepoints to the effective reproduction number given the serial
interval. The fact that it depends only on the observed incident counts makes it
reasonable to estimate \mathcal{R}_t . However, data collection idiosyncrasies can obscure this
relationship. Diagnostic testing targets symptomatic individuals, omitting
asymptomatic primary infections which can lead to future secondary infections. Testing
practices, availability, and uptake can vary across space and time [28, 29]. Finally,
incident cases as reported to public health are subject to delays due to laboratory
confirmation, test turnaround times, and eventual submission to public health [30]. For
these reasons, reported cases are lagging indicators of the course of the pandemic.
Furthermore, they do not represent the actual number of new infections that occur on a
given day, as indicated by exposure to the pathogen. The assumptions described above

(constant serial interval distribution, homogenous mixing, similar susceptibility and social behaviours, etc.) are therefore consequential. That said, Eq (2) also provides some comfort about deviations from these assumptions. If y_t is scaled by a constant (in time) describing the reporting ratio, then it will cancel from both sides. Similar arguments mean that even if such a scaling varies in time, as long as it varies slowly relative to the set of p_i that are larger than 0, Eq (2) will be a reasonably accurate approximation, so that \mathcal{R}_t can still be estimated well from reported incidence data. Finally, even a sudden change in reporting ratio, say from c_1 for $i = 1, \dots, t_1$ to c_2 for $i > t_1$ would only result in large errors for t in the neighbourhood of t_1 (where the size of this neighbourhood is again determined by the effective support of $\{p_i\}$). This robustness to certain types of data reporting issues partially justifies using Eq (2) to calculate \mathcal{R}_t . 163

2.2 Poisson trend filtering estimator 164

We use the daily confirmed incident cases y_t on day t to estimate the observed infectious cases under the model that y_t , given previous incident cases y_{t-1}, \dots, y_1 and a constant serial interval distribution, follows a Poisson distribution with mean Λ_t . That is, 165
166
167

$$y_t | y_1, \dots, y_{t-1} \sim \text{Poisson}(\Lambda_t), \text{ where } \Lambda_t = \mathcal{R}_t \sum_{i=1}^{t-1} p_i y_{t-i} = \mathcal{R}_t \eta_t. \quad (3)$$

Given a history of n confirmed incident counts $\mathbf{y} = (y_1, \dots, y_n)^\top$, our goal is to estimate \mathcal{R}_t . A natural approach is to maximize the likelihood, producing the maximum likelihood estimator (MLE): 168
169
170

$$\begin{aligned} \widehat{\mathcal{R}} &= \underset{\mathcal{R} \in \mathbb{R}_+^n}{\operatorname{argmax}} \mathbb{P}(\mathcal{R} | \mathbf{y}, \mathbf{p}) = \underset{\mathcal{R} \in \mathbb{R}_+^n}{\operatorname{argmax}} \prod_{t=1, \dots, n} \frac{(\mathcal{R}_t \eta_t)^{y_t} \exp\{-\mathcal{R}_t \eta_t\}}{y_t!} \\ &= \underset{\mathcal{R} \in \mathbb{R}_+^n}{\operatorname{argmin}} \sum_{t=1}^n \mathcal{R}_t \eta_t - y_t \log(\mathcal{R}_t \eta_t). \end{aligned} \quad (4)$$

This optimization problem, however, is easily seen to yield a one-to-one correspondence between the observations and the estimated effective reproduction number, i.e., $\widehat{\mathcal{R}}_t = y_t / \eta_t$, so that the estimated sequence $\widehat{\mathcal{R}}$ will have no significant smoothness. 171
172
173

The MLE is an unbiased estimator of the true parameter \mathcal{R}_t , but unfortunately has high variance: changes in y_t result in proportional changes in $\widehat{\mathcal{R}}_t$. To avoid this 174
175

behaviour, and to match the intuition that $\mathcal{R}_t \approx \mathcal{R}_{t-1}$, we advocate enforcing 176
smoothness of the effective reproduction numbers. This constraint will decrease the 177
estimation variance, and hopefully lead to more accurate estimation of \mathcal{R} , as long as the 178
smoothness assumption is reasonable. Smoothness assumptions are common (see 179
e.g., [16] or [1]), but the type of smoothness assumed is critical. [9] imposes smoothness 180
indirectly by estimating \mathcal{R}_t with moving windows of past observations. The Kalman 181
filter procedure of [16] would enforce in ℓ_2 -smoothness ($\int_0^n (\hat{\mathcal{R}}''(t))^2 dt < C$ for some C), 182
although the computational implementation results in $\hat{\mathcal{R}}$ taking values over a discrete 183
grid. [20] produces piecewise linear $\hat{\mathcal{R}}_t$, which turns out to be closely related to a special 184
case of our methodology. Smoother estimated curves will provide high-level information 185
about the entire epidemic, obscuring small local changes in $\mathcal{R}(t)$, but may also remove 186
the ability to detect large sudden changes, such as those resulting from lockdowns or 187
other major containment policies. 188

To enforce smoothness of $\hat{\mathcal{R}}_t$, we add a trend filtering penalty to Eq (5) [31–34]. 189
Because $\mathcal{R}_t > 0$, we explicitly penalize the divided differences (discrete derivatives) of 190
neighbouring values of $\log(\mathcal{R}_t)$. Let $\theta := \log(\mathcal{R}) \in \mathbb{R}^n$, so that $\Lambda_t = \eta_t \exp(\theta_t)$, and 191
 $\log(\eta_t \mathcal{R}_t) = \log(\eta_t) + \theta_t$. For evenly spaced incident case, we write our estimator as the 192
solution to the optimization problem 193

$$\hat{\mathcal{R}} = \exp(\hat{\theta}) \quad \text{where} \quad \hat{\theta} = \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \eta^\top \exp(\theta) - \mathbf{y}^\top \theta + \lambda \|D^{(k+1)}\theta\|_1, \quad (5)$$

where $\exp(\cdot)$ applies elementwise and $\|\mathbf{a}\|_1 := \sum_{i=1}^n |a_i|$ is the ℓ_1 norm. Here, 194
 $D^{(k+1)} \in \mathbb{Z}^{(n-k-1) \times n}$ is the $(k+1)^{\text{th}}$ order divided difference matrix for any 195
 $k \in \{0, \dots, n-1\}$. $D^{(1)} \in \{-1, 0, 1\}^{(n-1) \times n}$ is the divided difference matrix for $k=0$. 196
It is a sparse matrix with diagonal band of the form: 197

$$D^{(1)} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}. \quad (6)$$

$D^{(k+1)}$ for $k \geq 1$ is defined recursively as $D^{(k+1)} := D^{(1)} D^{(k)}$, where 198
 $D^{(1)} \in \{-1, 0, 1\}^{(n-k-1) \times (n-k)}$ takes the form defined in Eq (6). More description on 199

the recursive definition of divided difference matrix for trend filtering can be found
in [32, 33].

The tuning parameter (hyperparameter) λ balances data fidelity with desired
smoothness. When $\lambda = 0$, the problem in Eq (5) reduces to the MLE in Eq (4). Larger
tuning parameters privilege the regularization term and yield smoother estimates.
Finally, there exists λ_{\max} such that any $\lambda \geq \lambda_{\max}$ will result in $D^{(k+1)}\hat{\theta} = 0$ and $\hat{\theta}$ will
be the Kullback-Leibler projection of \mathbf{y} onto the null space of $D^{(k+1)}$ (see
subsection 2.3).

The solution to Eq (5) will result in piecewise polynomials, specifically called
discrete splines. For example, 0th-degree discrete splines are piecewise constant,
1st-degree curves are piecewise linear, and 2nd-degree curves are piecewise quadratic.
For $k \geq 1$, k th-degree discrete splines are continuous and have continuous discrete
differences up to degree $k - 1$ at the knots. This penalty results in more flexibility
compared to the homogeneous smoothness that is created by the squared ℓ_2 norm.
Using different orders of divided differences result in estimated effective reproduction
numbers with different smoothness properties.

For unevenly-spaced data, the spacing between neighbouring parameters varies with
the time between observations, and thus, the divided differences must be adjusted by
the times that the observations occur. Given observation times $\mathbf{x} = (x_1, \dots, x_n)^T$, for
 $k \geq 1$, define a k th-order diagonal matrix

$$X^{(k)} = \text{diag} \left(\frac{k}{x_{k+1} - x_1}, \frac{k}{x_{k+2} - x_2}, \dots, \frac{k}{x_n - x_{n-k}} \right). \quad (7)$$

Letting $D^{(\mathbf{x}, 1)} := D^{(1)}$, then for $k \geq 1$, the $(k + 1)$ th-order divided difference matrix for
unevenly spaced data can be created recursively by $D^{(\mathbf{x}, k+1)} := D^{(1)} X^{(k)} D^{(\mathbf{x}, k)}$. No
adjustment is required for $k = 0$.

Due to the penalty structure, this estimator is locally adaptive, meaning that it can
potentially capture local changes such as the initiation of control measures. [19, 20]
considered only the 2nd-order divided difference of \mathcal{R}_t rather than its logarithm. In
comparison to their work, our estimator (i) allows for arbitrary degrees of temporal
smoothness and (ii) avoids the potential numerical issues of penalizing/estimating
positive real values. Furthermore, as we will describe below, our procedure is

computationally efficient for estimation over an entire sequence of penalty strengths λ 229
and provides methods for choosing how smooth the final estimate should be. 230

2.3 Solving over a sequence of tuning parameters 231

We can solve the Poisson trend filtering estimator over an arbitrary sequence of λ that 232
produces different levels of smoothness in the estimated curves. We consider a 233
candidate set of M λ -values, $\boldsymbol{\lambda} = \{\lambda_m\}_{m=1}^M$, that is strictly decreasing. 234

Let $D := D^{(k+1)}$ for simplicity in the remainder of this section. As $\lambda \rightarrow \infty$, the 235
penalty term $\lambda \|D\theta\|_1$ dominants the Poisson objective, so that minimizing the objective 236
is asymptotically equivalent to minimizing the penalty term, which results in $\|D\theta\|_1 = 0$. 237
In this case, the divided differences of θ with order $k + 1$ is always 0, and thus, θ must 238
lie in the null space of D , that is, $\theta \in \mathcal{N}(D)$. The same happens for any λ beyond this 239
threshold, so define λ_{\max} to be the smallest λ that produces $\theta \in \mathcal{N}(D)$. It turns out 240
that this value can be written explicitly as $\lambda_{\max} = \|(D^\dagger)^\top (\eta - y)\|_\infty$, where D^\dagger is the 241
(left) generalized inverse of D satisfying $D^\dagger D = I$ and $\|a\|_\infty := \max_{i=1}^n \{|a_i|\}$ is the 242
infinity norm. Therefore, we use $\lambda_1 = \lambda_{\max}$ and then choose the minimum λ_M to be 243
 $r\lambda_{\max}$ for some $r \in (0, 1)$ (typically $r = 10^{-5}$). Given any $M \geq 3$, we generate a 244
sequence of λ values to be equally spaced on the log-scale between λ_1 and λ_M . 245

To compute the sequence efficiently, the model is estimated sequentially by visiting 246
each component of $\boldsymbol{\lambda}$ in order. The estimates produced for a larger λ are used as the 247
initial values (warm starts) for the next smaller λ . By solving through the entire 248
sequence of tuning parameters, we have a better chance to achieve a better trade-off 249
between bias and variance, and accordingly, improved accuracy relative to procedures 250
examining one fixed value of λ . 251

2.4 Choosing a final λ 252

We estimate model accuracy over the candidate set through K -fold cross validation 253
(CV) to choose the best tuning parameter. Specifically, we divide \mathbf{y} (except the first and 254
last observations) roughly evenly and randomly into K folds, estimate \mathcal{R}_t for all $\boldsymbol{\lambda}$ 255
leaving one fold out, and then predict the held-out observations. Model accuracy can be 256
measured by multiple metrics such as mean squared error $\text{MSE}(\hat{y}, y) = n^{-1} \|\hat{y} - y\|_2^2$ or 257

mean absolute error $\text{MAE}(\hat{y}, y) = n^{-1} \|\hat{y} - y\|_1$, but we prefer to use the (average) deviance, to mimic the likelihood in Eq (4):

$D(y, \hat{y}) = n^{-1} \sum_{i=1}^n 2(y_i \log(y_i) - y_i \log(\hat{y}_i) - y_i + \hat{y}_i)$, with the convention that $0 \log(0) = 0$. Note that for any K and any M , we will end up estimating the model $(K + 1)M$ times rather than once.

2.5 Approximate confidence bands

We also provide empirical confidence bands of the estimators with approximate coverage. Consider the related estimator $\tilde{\mathcal{R}}_t$ defined as

$$\tilde{\mathcal{R}} = \exp(\tilde{\theta}) \quad \text{where} \quad \tilde{\theta} = \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \eta^\top \exp(\theta) - \mathbf{y}^\top \theta + \lambda \|D\theta\|_2^2. \quad (8)$$

Let $\tilde{\mathbf{y}} = \eta \circ \tilde{\mathcal{R}}$, and then it can be shown (for example, Theorem 2 in [35]) that an estimator for $\text{Var}(\tilde{\mathbf{y}})$ is given by $(\text{diag}(\tilde{\mathbf{y}}^{-2}) + \lambda D^\top D)^\dagger$. Finally, an application of the delta method shows that $\text{Var}(\tilde{\mathbf{y}}_t)/\eta_t^2$ is an estimator for $\text{Var}(\tilde{\mathcal{R}}_t)$ for each $t = 1, \dots, n$. We therefore use $(\text{diag}(\tilde{\mathbf{y}}^{-2}) + \lambda D^\top D)_t^\dagger / \eta_t^2$ as an estimator for $\text{Var}(\widehat{\mathcal{R}}_t)$. An approximate $(1 - \alpha)\%$ confidence interval then can be written as $\widehat{\mathcal{R}}_t \pm s_t \times T_{\alpha/2, n-\text{df}}$, where s_t is the square-root of $\text{Var}(\widehat{\mathcal{R}}_t)$ for each $t = 1, \dots, n$ and df is the number of changepoints in $\widehat{\theta}$ plus $k + 1$.

2.6 Bayesian perspective

Unlike many other methods for \mathcal{R}_t estimation, our approach is frequentist rather than Bayesian. Nonetheless, it has a corresponding Bayesian interpretation: as a state-space model with Poisson observational noise, autoregressive transition equation of degree $k \geq 0$, e.g., $\theta_{t+1} = 2\theta_t - \theta_{t-1} + \varepsilon_{t+1}$ for $k = 1$, and Laplace transition noise $\varepsilon_{t+1} \sim \text{Laplace}(0, 1/\lambda)$. Compared to **EpiFilter** [16], we share the same observational assumptions, but our approach has a different transition noise. **EpiFilter** estimates the posterior distribution of \mathcal{R}_t , and thus it can provide credible interval estimates as well. Our approach produces the maximum *a posteriori* estimate via an efficient convex optimization, obviating the need for MCMC sampling. But the associated confidence bands are created differently.

3 Results

Implementation of our approach is provided in the R package `rtestim`. All computational experiments are conducted on the Cedar cluster provided by Compute Canada with R 4.3.1. The R packages used for simulation and real-data application are `EpiEstim` 2.2-4, `EpiLPS` 1.2.0, and `rtestim` 0.0.4.

3.1 Synthetic experiments

We simulate four scenarios of the time-varying effective reproduction number, intended to mimic different epidemics. The first two scenarios are rapidly controlled by intervention, where the $\mathcal{R}(t)$ consists of one discontinuity and two segments. Scenario 1 has constant $\mathcal{R}(t)$ before and after an intervention, while Scenario 2 grows exponentially, then decays. The other two scenarios are more complicated, where more waves are involved. Scenario 3 has four linear segments with three discontinuities, which reflect the effect of an intervention, resurgence to rapid transmission, and finally suppression of the epidemic. Scenario 4 involves sinusoidal waves throughout the epidemic. The first three scenarios and the last scenario are motivated by [16] and [17] respectively. We name the four scenarios as (1) *piecewise constant*, (2) *piecewise exponential*, (3) *piecewise linear*, and (4) *periodic* lines or curves respectively.

In all cases, the times of observation are regular, and epidemics are of length $n = 300$. Specifically, in Scenario 1, $\mathcal{R}_t = 2, 0.8$ before and after $t = 120$. In Scenario 2, \mathcal{R}_t increases and decreases exponentially with rates 0.01, 0.005 pre and post $t = 100$. In Scenario 3, \mathcal{R}_t is piecewise linear with four discontinuous segments following

$$\begin{aligned} \mathcal{R}(t) &= \left(2.5 - \frac{0.5}{74}(t-1)\right) \mathbf{1}_{[1,76)}(t) + \left(0.8 - \frac{0.2}{74}(t-76)\right) \mathbf{1}_{[76,151)}(t) \\ &\quad + \left(1.7 + \frac{0.3}{74}(t-151)\right) \mathbf{1}_{[151,226)}(t) + \left(0.9 - \frac{0.4}{74}(t-226)\right) \mathbf{1}_{[226,300]}(t), \end{aligned} \quad (9)$$

where $\mathbf{1}_A(t) = 1$, if $t \in A$, and $\mathbf{1}_A(t) = 0$ otherwise. In Scenario 4, \mathcal{R}_t is realization of the continuous, periodic curve generated by the function

$$\mathcal{R}(t) = 0.2((\sin(\pi t/12) + 1) + (2 \sin(5\pi t/12) + 2) + (3 \sin(5\pi t/6) + 3)), \quad (10)$$

evaluated at equally spaced points $t \in [0, 10]$. These \mathcal{R}_t scenarios are illustrated in Fig 2. 307

We use serial interval (SI) distributions of measles (with mean 14.9 and standard 308 deviation 3.9) at Hagelloch, Germany in 1861 [36] and SARS (with mean 8.4 and 309 standard deviation 3.8) at Hong Kong in 2003 [37], inspired by [10], to generate 310 synthetic epidemics. We initialize all epidemics with $y_1 = 2$ cases and generate for 311 $t = 2, \dots, 300$. We compute the expected incidence Λ_t using the renewal equation, and 312 generate the incident infections from the Poisson distribution $y_t \sim \text{Pois}(\Lambda_t)$. The 313 synthetic measles epidemics tend to have smaller incident cases, and the SARS 314 epidemics tend to have larger incidence. The intuition behind this is since the mean of 315 serial interval of SARS is smaller with a similar standard deviation compared to the 316 counterpart of measles, SARS epidemics have an averaged shorter period between 317 primary and secondary onsets of symptoms of the infected individuals, and then they 318 can result in a larger increase in incidence within the same period of time. To verify the 319 performance of our model under the violation of this distributional assumption, we also 320 generate incident infections using the negative Binomial distribution with dispersion size 321 5, i.e., $y_t \sim \text{NB}(\text{mean} = \Lambda_t, \text{size} = 5)$, which gives a reasonably large overdispersion. 322 For each problem setting (including a SI distribution, a \mathcal{R}_t scenario, and an incidence 323 distribution), we generate 50 random samples, resulting in 800 total synthetic epidemics. 324 An example of measles and SARS epidemics for each effective reproduction number 325 scenario with an incidence distribution is displayed in Fig 2. We also visualize the 326 (over)dispersion level of the following synthetic epidemics in Appendix. 327

We compare `RtEstim` to `EpiEstim`, `EpiLPS`, and `EpiFilter`. `EpiEstim` estimates 328 the posterior distribution of the effective reproduction number given a Gamma prior 329 and Poisson distributed observations over a trailing window, under the assumption that 330 the effective reproduction number is constant during that window. A larger window 331 averages out more fluctuations, leading to smoother estimates, whereas, a shorter 332 sliding window is more responsive to sudden spikes or declines. We tried the weekly 333 sliding window, as well as a monthly window. However, since neither considerably 334 outperforms the other across all scenarios, we defer the monthly results to the 335 supplementary document. `EpiLPS` is another Bayesian approach that estimates 336 P-splines based on the Laplace approximation to the conditional posterior with negative 337 Binomial likelihood. `EpiFilter` is also a Bayesian approach that smooths \mathcal{R}_t at each 338

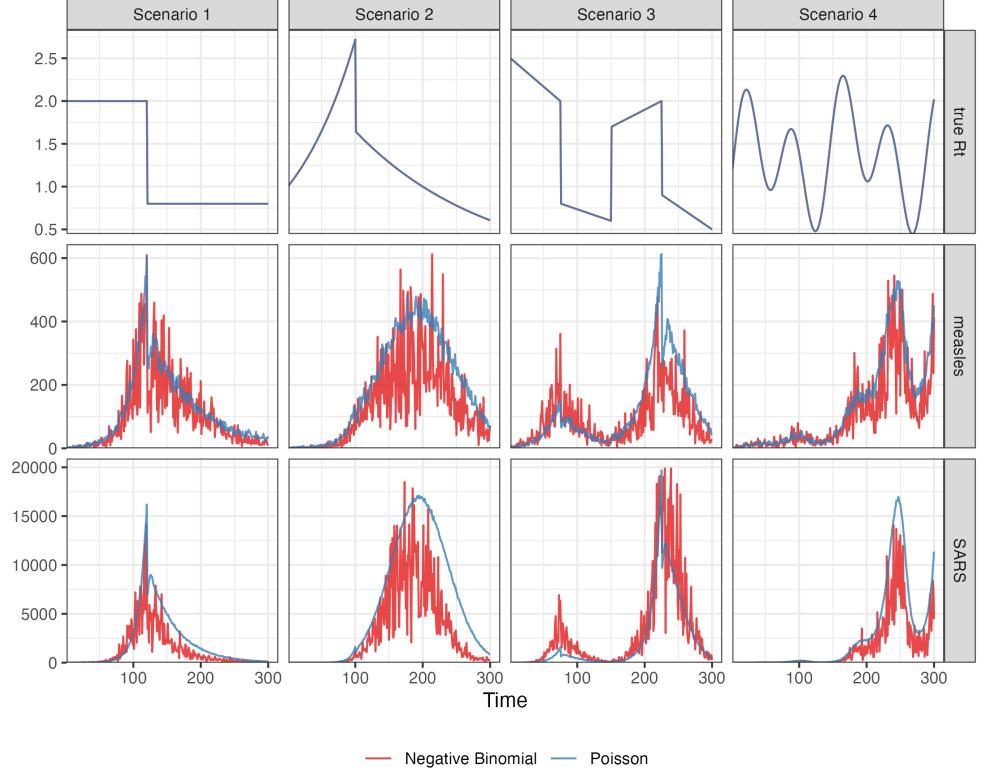


Fig 2. The effective reproduction numbers for four \mathcal{R}_t scenarios (*in the top row*). The sample measles (*in the middle row*) and SARS (*in the bottom row*) incident cases drawn from Poisson (*in blue curves*) or negative Binomial (*in red curves*) distribution across 4 \mathcal{R}_t scenarios (*in four columns respectively*).

timepoint given all observed incidence, improved upon the filtering methods that filter 339
 \mathcal{R}_t given the observations prior to and on time t . We apply **RtEstim** with four degrees, 340
(1) piecewise constant $k = 0$, (2) piecewise linear & cubic $k = 1, 3$, (3) piecewise linear 341
 $k = 1$ and (4) piecewise cubic polynomials $k = 3$, to solve all settings. In each case, we 342
examine a grid of 50 λ values, selecting the best using 10-fold cross validation. For all 343
models and problems, we use the same serial interval distribution for estimation that 344
was used to create the data. Taking different hyper parameters into consideration, we 345
solve each problem using 7 methods in total. 346

Throughout the four \mathcal{R}_t scenarios, the degrees of **RtEstim** can be correctly or 347
wrongly specified. On one hand, our method can take the advantage of a correctly 348
specified degree of piecewise polynomials compared to other methods, since the 349
competitors only consider one fixed degree of smoothness and do not allow a 350
user-specified degree. (We will discuss more in Section 4.) On the other hand, we will 351

illustrate that a wrongly specified degree can still result in relatively accurate \mathcal{R}_t
estimates in the following experimental results.

To measure estimation accuracy, we compare $\widehat{\mathcal{R}}$ to \mathcal{R} using the Kullback-Leibler
(KL) divergence. We use the KL divergence for the Poisson distribution (summed over
across all t) to measure the accuracy of the \mathcal{R}_t estimates

$$D_{KL}(\mathcal{R} \parallel \widehat{\mathcal{R}}) = \sum_{t=1}^n w_t \left(\mathcal{R}_t \log \left(\frac{\mathcal{R}_t}{\widehat{\mathcal{R}}_t} \right) + \widehat{\mathcal{R}}_t - \mathcal{R}_t \right), \quad (11)$$

where $\mathcal{R} = \{\mathcal{R}_t\}_{t=1}^n$ and $w_t = \eta_t / \sum_t \eta_t$ is the rescaled total infectiousness. To fairly
compare across methods, we drop the estimates during the first week because estimates
from `EpiEstim` do not begin until $t = 8$ (using a weekly window). KL divergence is
more appropriate for measuring accuracy because it connects directly to the Poisson
likelihood used to generate the data, whereas standard measures like the mean-squared
error correspond to Gaussian likelihood. Using Poisson likelihood has the effect of
increasing the relative cost of mistakes when Λ_t is small. Other details of the
experimental settings are deferred to the supplementary document.

3.2 Results for synthetic data

`RtEstim` overall outperforms `EpiEstim` and `EpiLPS` in the experimental study. [Fig 3](#)
and [Fig 4](#) visualizes the KL divergence across the seven models. Under both Poisson and
negative Binomial distributions, `RtEstim` is easily the most accurate for Scenarios 1 and
3: the median of KL divergence is much lower and the boxes frequently fail to overlap
indicating better performance than the other two methods across all 50 simulations. The
advantage is less pronounced for the negative Binomial configuration, but still obvious.
`RtEstim` and `EpiLPS` have similar performance in Scenarios 2 and 4. For the Poisson
case, `RtEstim` and `EpiLPS` both have very small KL scores, which are very close to zero.
In Scenario 4, `RtEstim` is slightly better for Poisson and `EpiLPS` is better for negative
Binomial, but the boxes largely overlap each other. `EpiLPS` has a slightly lower median
and a smaller IQR in Scenario 2 for the negative Binomial case. Both smoothness
choices for `RtEstim` in Scenario 2 perform similarly across noise distributions, implying
good performance under model misspecification. We will examine a single realization of
each experiment to investigate these global conclusions in more detail.

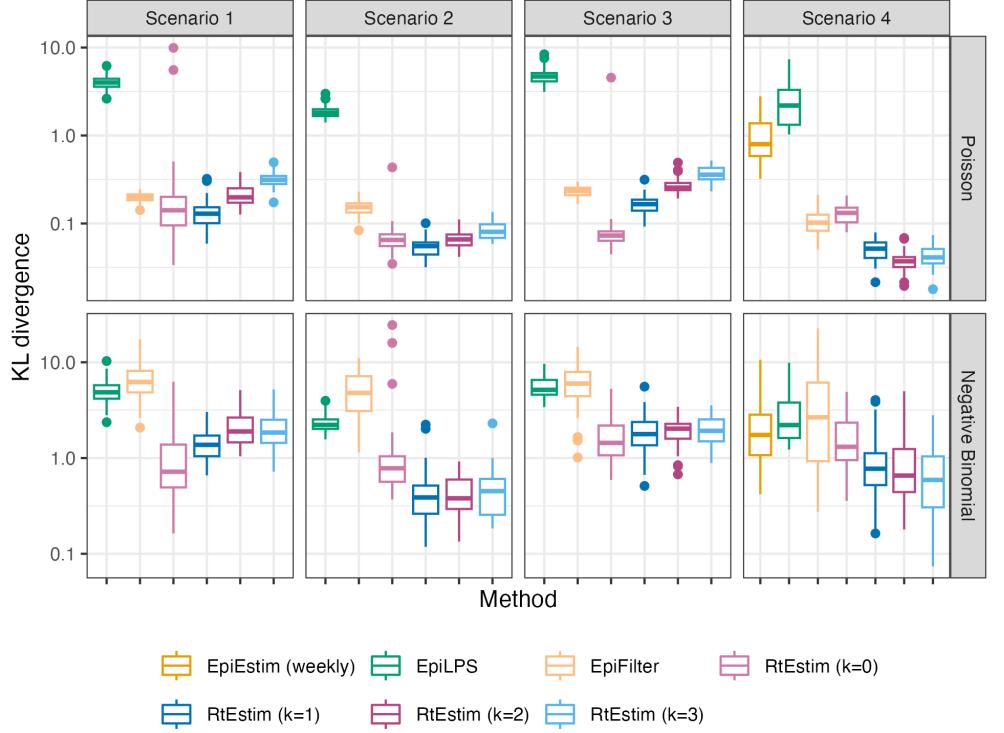


Fig 3. Boxplot of KL divergence between the estimated \hat{R}_t and the true R_t across 50 synthetic measles epidemics for each approach given Poisson incidence (*in top panels*) and negative Binomial incidence (*in bottom panels*) respectively. Outliers are excluded.

Fig 5 show one realization for the estimated effective reproduction number under the Poisson generative model for all four scenarios. Compared to EpiEstim and EpiLPS, which have rather severe difficulties at the beginning of the time series, RtEstim estimates are more accurate—they nearly overlap with the true values—without suffering from the initialization problem. Scenario 1 is the simplest case with only one knot and two constant segments. Besides the edge problem, EpiEstim and EpiLPS produce “smooth” estimated curves that are continuous at the changepoint, which results in large mistakes in that neighbourhood. Since the piecewise constant RtEstim estimator does not force any smoothness in R_t , it easily captures the sharp change. Scenario 2 is relatively easy for all methods, except at the changepoint occurring at the end of the exponential growth. Although the truth is likely best represented with a discontinuous piecewise cubic curve, the actual curvature is so gentle that linear estimation ($k = 1$) appears potentially reasonable. However, RtEstim has difficulty recovering the acute rise in the growth phase because it enforces continuity at the

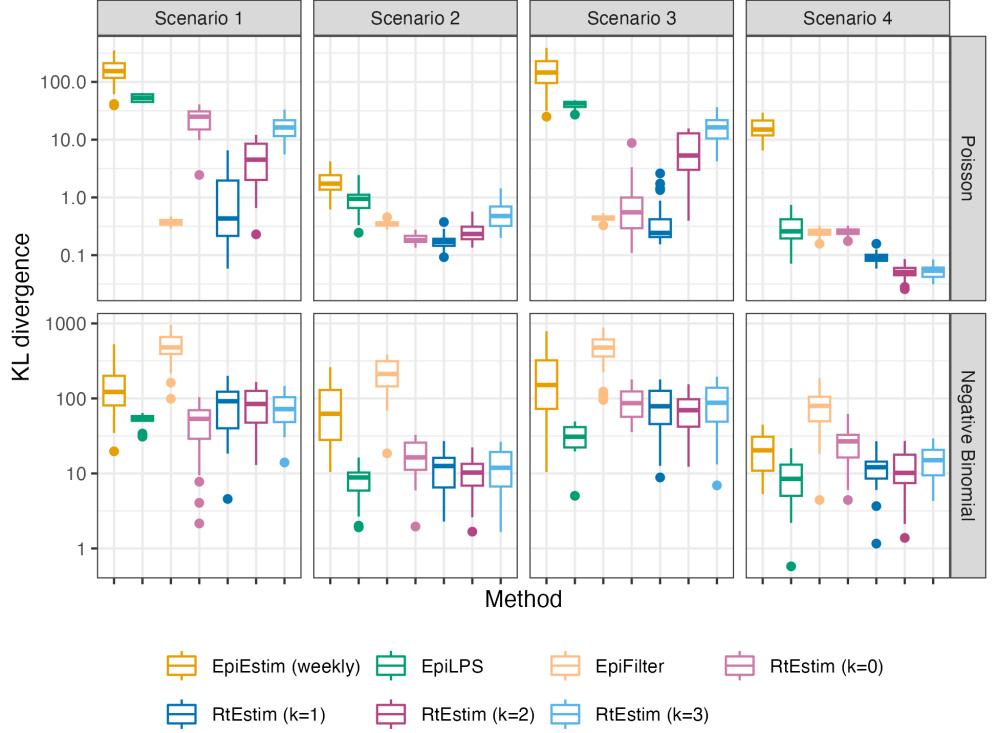


Fig 4. Boxplot of KL divergence between the estimated \hat{R}_t and the true R_t across 50 synthetic SARS epidemics for each approach given Poisson incidence (*in top panels*) and negative Binomial incidence (*in bottom panels*) respectively. Outliers are excluded.

change point.

To investigate the performance when the Poisson assumption (imposed by both `RtEstim` and `EpiEstim`) is violated, we also examine estimation accuracy with negative Binomial data. Fig 6 displays a realization, analogous to the previous case, for all methods and scenarios. `RtEstim` has more difficulty relative to the Poisson setting, especially at the beginning of the outbreak. This is most pronounced in Scenario 4, where `RtEstim` is overly smooth, except in the last wave. In Scenario 2, `RtEstim` successfully captures the change point, but suffers from the same discontinuity problem as in the Poisson setting. In Scenario 3, the piecewise linear version of `RtEstim` recovers the curvature of R_t well, but is less accurate than in the Poisson case.

Finally, it is important to provide a brief comparison of the running times of all three models across the 8 experimental settings. We find that almost all models across all experiments complete within 10 seconds. `RtEstim` generally takes the longest, due to a relatively large number of estimates—50 values of λ and 10 folds of cross validation

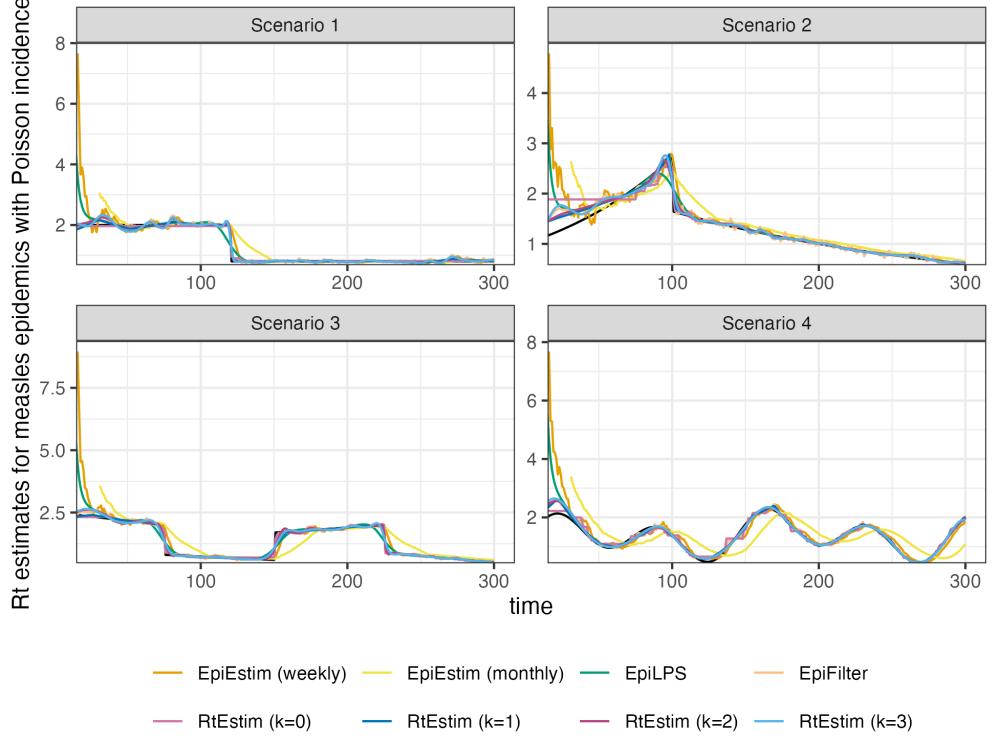


Fig 5. Example of effective reproduction number estimation for measles epidemics with Poisson observations.

require 550 estimates—while other models run only a single time for a fixed setting of hyperparameters per experiment. Additional results on timing comparisons are deferred to the supplementary document.

3.3 Real-data results: Covid-19 incident cases in British Columbia

We implement `RtEstim` on Covid-19 confirmed incident cases in British Columbia (B.C.) as reported on May 18, 2023 (visualized in Fig 7) by the B.C. Centre for Disease Control [38]. We use the gamma distribution with shape 2.5 and scale 2.5 to approximate the serial interval function, which is similar to empirical estimates [39].

Considering the first, second, and third polynomial degrees, $\hat{\mathcal{R}}_t$ for Covid-19 in British Columbia (illustrated in Fig 8) is always less than 3 except at the very early stage, which means that one distinct infected individuals on average infects less than three other individuals in the population. Examining three different settings for k , the temporal evolution of $\hat{\mathcal{R}}$ (across all regularization levels λ) are similar near the highest

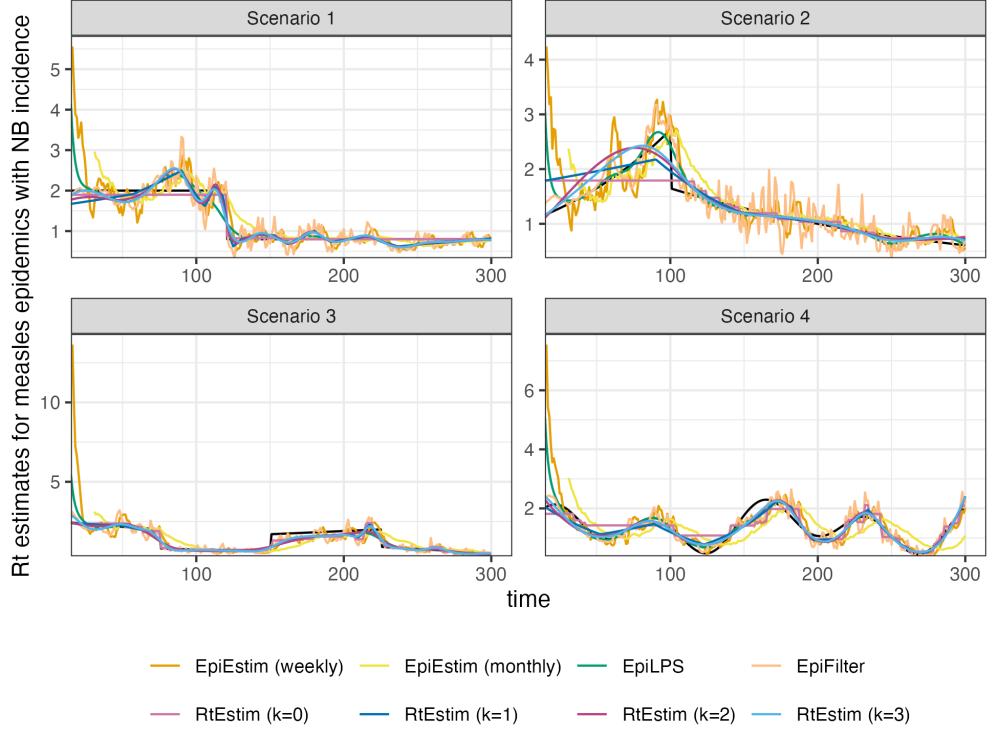


Fig 6. Example of effective reproduction number estimation for measles epidemics with negative Binomial observations.

peak around the end of 2021 before dropping shortly thereafter. Throughout the
 estimated curves, the peaks and troughs of the effective reproduction numbers precede
 the growth and decay cycles of confirmed cases, as expected. We also visualize 95%
 confidence bands for the point estimates with λ chosen by minimizing cross-validated
 KL divergence in Fig 8.
 422
 423
 424
 425
 426

The estimated effective reproduction numbers are relatively unstable before April,
 2022. The highest peak coincides with the emergence and global spread of the Omicron
 variant. The estimated effective reproduction numbers fall below 1 during two time
 periods—roughly from April, 2021 to July, 2021 and from January, 2022 to April, 2022.
 The first trough coincides with the introduction of Covid-19 vaccines in British
 Columbia. The second trough, shortly after the largest peak may be due to variety of
 factors resulting in the depletion of the susceptible population such as increased
 self-isolation in response to media coverage of the peak or immunity incurred via recent
 infection. Since April, 2022, the estimated effective reproduction number has remained
 relatively stable (fluctuating around one) corresponding to low reported cases, though
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436

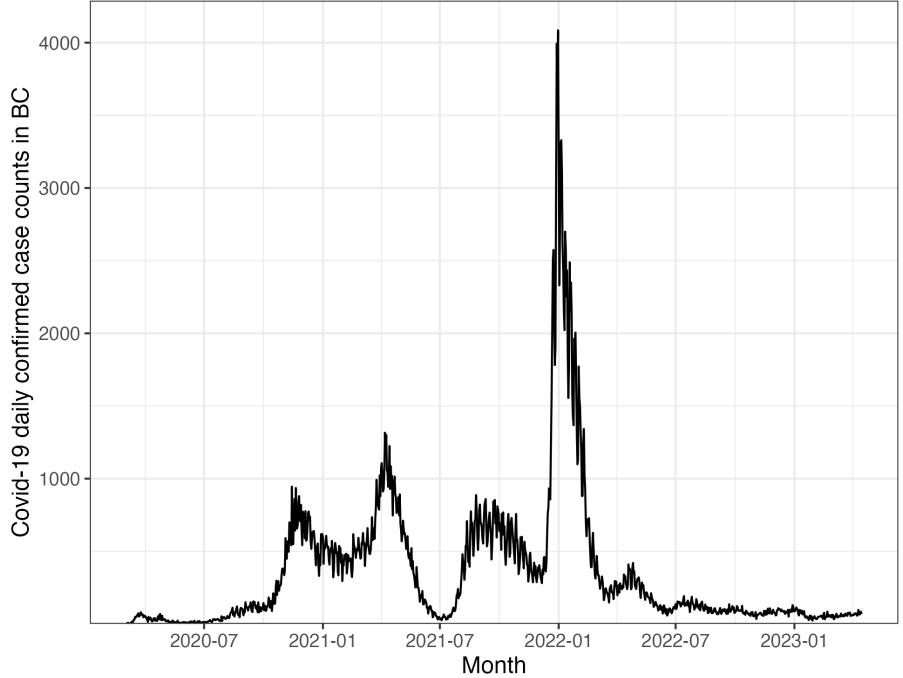


Fig 7. Covid-19 daily confirmed incident cases between March 1st, 2020 and April 15th, 2023 in British Columbia, Canada.

reporting behaviours also changed significantly since the Omicron wave. 437

3.4 Real-data results: influenza in Baltimore, Maryland, 1918 438

We also apply `RtEstim` to daily reported influenza cases in Baltimore, Maryland 439 occurring during the world-wide pandemic of 1918 from September to November [40]. 440 The data, shown in Fig 9, is included in the `EpiEstim` R package. The 1918 influenza 441 outbreak, caused by the H1N1 influenza A virus, was unprecedentedly deadly with case 442 fatality rate over 2.5%, infecting almost one-third of the population across the 443 world [41]. The CV-tuned piecewise cubic estimates in Fig 10 better capture the growth 444 at the beginning of the pandemic in Fig 9. The estimated \mathcal{R}_t curve suggests that the 445 transmissibility of the pandemic grew rapidly over the first 30 days before declining 446 below one after 50 days. However, it also suggests an increase in infectiousness toward 447 the end of the period. With this data, it is difficult to determine if there is a second 448 wave or a steady decline ahead. The CV-tuned piecewise constant and linear estimates 449 in Fig 10 both suggest a steady decline. This conclusion is supported by the fact that 450 incident cases decline to zero at the end of the period and matches \mathcal{R} estimates in [10], 451

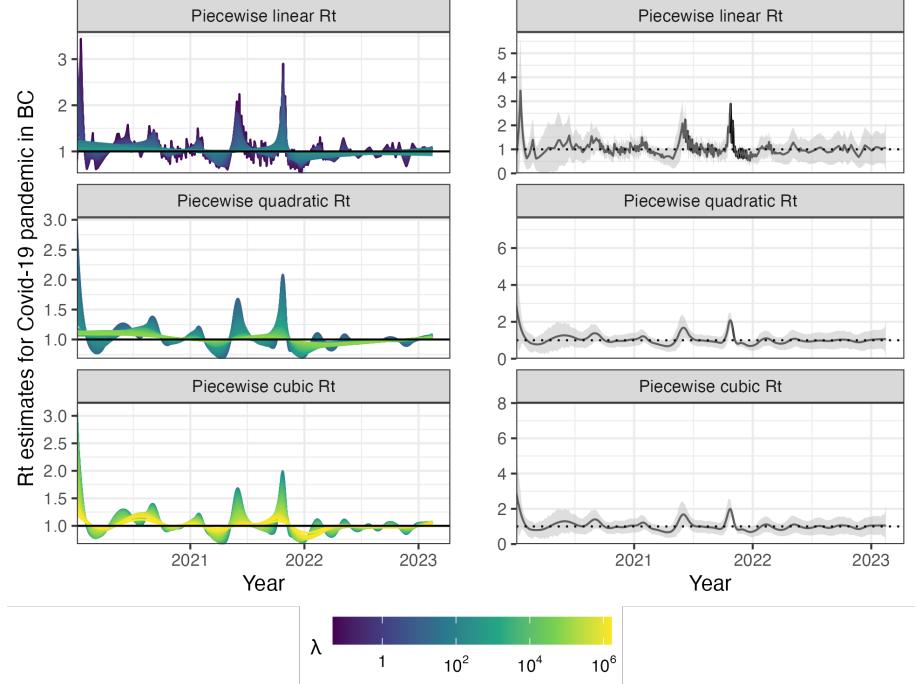


Fig 8. Estimated effective reproduction number based on Covid-19 daily confirmed incident cases between March 1st, 2020 and April 15th, 2023 in British Columbia, Canada. The left panels show estimates corresponding to 50 tuning parameters. The right panels show the CV-tuned estimate along with approximate 95% confidence bands. The top, middle and bottom panels show the estimated \mathcal{R}_t using the Poisson trend filtering in Eq (5) with degrees $k = 1, 2, 3$ respectively.

which are all lower than one. 452

4 Discussion 453

The **RtEstim** methodology provides a locally adaptive estimator using Poisson trend filtering on univariate data. It captures the heterogeneous smoothness of effective reproduction numbers given observed incidence data rather than resulting in global smoothness. This is a nonparametric regression model which can be written as a convex optimization (minimization) problem. Minimizing the distance (KL divergence across all coordinates) between the estimators and (functions of) observations guarantees data fidelity while the penalty on divided differences between pairs of neighbouring parameters imposes smoothness. The ℓ_1 -regularization results in sparsity of the divided differences, which leads to heterogeneous smoothness across time. 454
455
456
457
458
459
460
461
462

The property of local adaptivity (heterogenous smoothness) is useful to 463

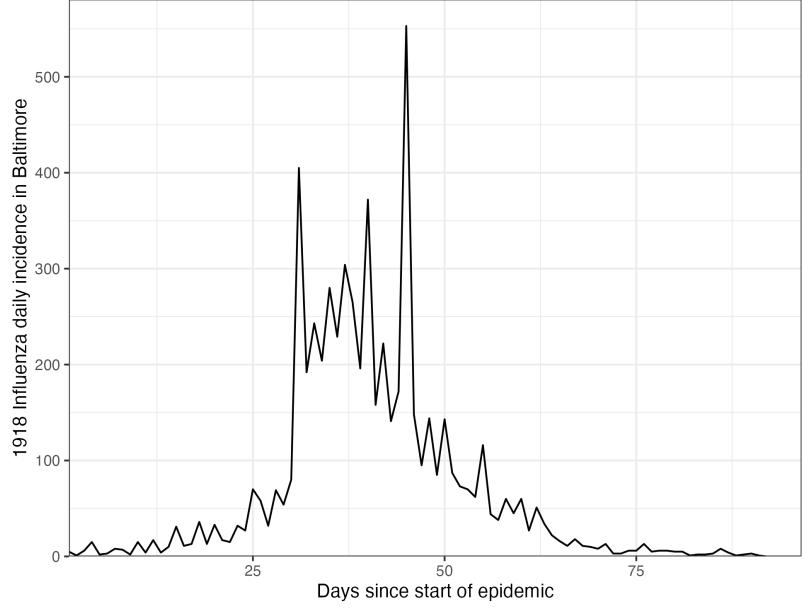


Fig 9. Daily incident influenza cases in Baltimore, Maryland between September and November in 1918.

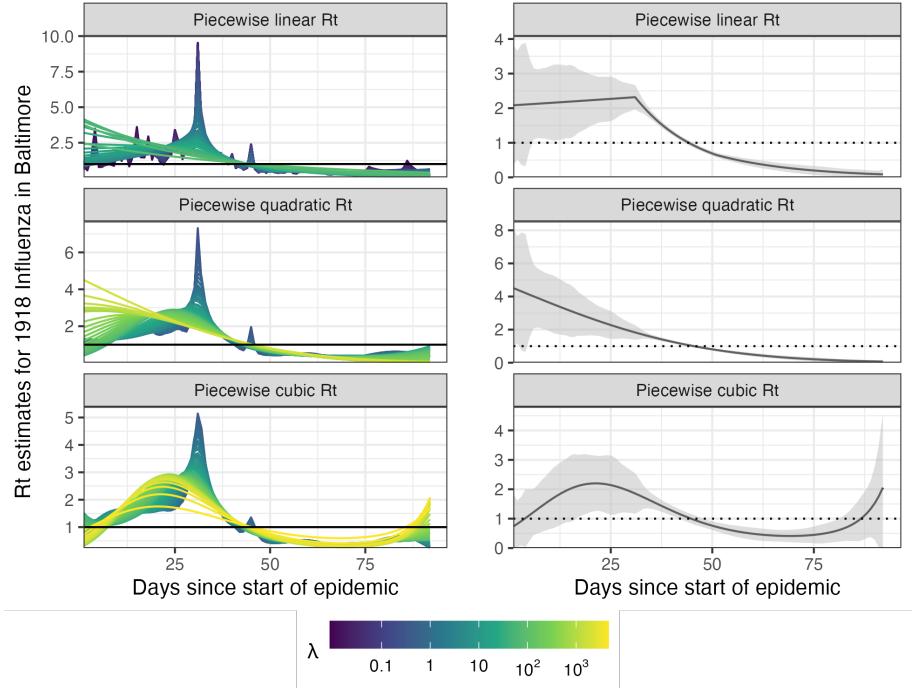


Fig 10. Estimated effective reproduction numbers for influenza in Baltimore, Maryland in 1918. The left panels show estimates for a set of 50 tuning parameters. The right column displays the CV-tuned estimates with approximate 95% confidence bands. The rows (top to bottom) show estimated effective reproduction numbers (\mathcal{R}_t) using the Poisson trend filtering in Eq (5) with $k = 1, 2, 3$ respectively.

automatically distinguish, for example, seasonal outbreaks from outbreaks driven by
other factors (behavioural changes, foreign introduction, etc.). Given a well-chosen
polynomial degree, the growth rates can be quickly detected, potentially advising public
health authorities to implement policy changes. The effective reproduction numbers can
be estimated retrospectively to examine the efficacy of such policies, whether they result
in \mathcal{R}_t falling below 1 or the speed of their effects. The smoothness of \mathcal{R}_t curves
(including the polynomial degrees and tuning parameters) should be chosen based on
the purpose of the study in practice.

Our method **RtEstim** provides a natural way to deal with missing data, for example,
on weekends and holidays or due to changes in reporting frequency. While solving the
convex optimization problem, our method can easily handle uneven spacing or irregular
reporting. Computing the total primary infectiousness is also easily generalized to
irregular reporting by modifying the discretization of the serial interval distribution.
Additionally, because the ℓ_1 penalty introduces sparsity (operating like a median rather
than a mean), this procedure is relatively insensitive to outliers compared to ℓ_2
regularization.

There are a number of limitations that may influence the quality of \mathcal{R}_t estimation.
While our model is generic for incidence data rather than tailored to any specific
disease, it does assume that the generation interval is short relative to the period of
data collection. More specialized methodologies would be required for diseases with long
incubation periods such as HIV or Hepatitis. Our approach, does not explicitly model
imported cases, nor distinguish between subpopulations that may have different mixing
behaviour. While the Poisson assumption is common, it does not handle overdispersion
(observation variance larger than the mean). The negative binomial distribution is a
good alternative, but more difficult to estimate in this context. As described in
[section 1](#), the expression for \mathcal{R} assumes that a relatively constant proportion of true
infections is reported. However, if this proportion varies with time (say, due to changes
in surveillance practices or testing recommendations), the estimates may be biased over
this window. A good example is in early January 2022, during the height of the
Omicron wave, British Columbia moved from testing all symptomatic individuals to
testing only those in at-risk groups. The result was a sudden change that would render
 \mathcal{R}_t estimates on either side of this timepoint incommensurable.

As currently implemented, **RtEstim** uses a fixed serial interval throughout the period of study, but as factors such as population immunity vary, the serial interval may vary as well [12]. Another issue relates to equating serial and generation intervals (also mentioned above). The serial interval distribution is generally wider than that of the generation interval, because the serial interval involves the convolution of two distributions, and is unlikely to actually follow a named distribution like Gamma, though it may be reasonably well approximated by one. Our implementation allows for an arbitrary distribution to be used, but requires the user to specify the discretization explicitly, requiring more nuanced knowledge than is typically available. Pushing this analysis further, to accommodate other types of incidence data (hospitalizations or deaths), a modified generation interval distribution would be necessary, and further assumptions would be required as well. Or else, one would first need to deconvolve deaths to infection onset before using our software.

Nonetheless, our methodology is implemented in a lightweight R package **rtestim** and computed efficiently, especially for large-scale data, with a proximal Newton solver coded in C++. Given available incident case data, prespecified serial interval distribution, and a choice of degree k , **RtEstim** is able to produce accurate estimates of effective reproduction number and provide efficient tuning parameter selection via cross validation.

Acknowledgments

This research was enabled in part by support provided by BC DRI group who manages Cedar cloud (<https://docs.alliancecan.ca/wiki/Cedar>) and the Digital Research Alliance of Canada (alliancecan.ca).

References

1. Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. Practical considerations for measuring the effective reproductive number, R_t . PLoS Computational Biology. 2020;16(12):e1008409.

2. Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. Mathematical and statistical estimation approaches in epidemiology. 2009; p. 103–121.
3. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. *PloS one*. 2007;2(8):e758.
4. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford university press; 1991.
5. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*. 2004;160(6):509–516.
6. Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature*. 2020;584(7821):420–424.
7. Goldstein IH, Parker DM, Jiang S, Minin VM. Semiparametric inference of effective reproduction number dynamics from wastewater pathogen surveillance data. *arXiv preprint arXiv:230815770*. 2023;.
8. Goldstein IH, Wakefield J, Minin VN. Incorporating testing volume into estimation of effective reproduction number dynamics. *arXiv preprint arXiv:220804418*. 2022;.
9. Cori A, Cauchemez S, Ferguson NM, Fraser C, Dahlqwist E, Demarsh PA, et al.. EpiEstim: estimate time varying reproduction numbers from epidemic curves; 2020.
10. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*. 2013;178(9):1505–1512.
11. Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*. 2019;29:100356.

12. Nash RK, Bhatt S, Cori A, Nouvellet P. Estimating the epidemic reproduction number from temporally aggregated incidence data: A statistical modelling approach and software tool. *PLoS Computational Biology*. 2023;19(8):e1011439.
13. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*. 2020;5:112.
14. Abbott S, Funk S, Hickson J, Badr HS, Monticone P, Ellis P, et al.. *epiforecasts/EpiNow2: 1.4.0 release*; 2023.
15. Lison A, Abbott S, Huisman J, Stadler T. Generative Bayesian modeling to nowcast the effective reproduction number from line list data with missing symptom onset dates. *arXiv preprint arXiv:230813262*. 2023;.
16. Parag KV. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Computational Biology*. 2021;17(9):e1009347.
17. Gressani O, Wallinga J, Althaus CL, Hens N, Faes C. EpiLPS: A fast and flexible Bayesian tool for estimation of the time-varying reproduction number. *PLoS Computational Biology*. 2022;18(10):e1010618.
18. Trevisin C, Bertuzzo E, Pasetto D, Mari L, Miccoli S, Casagrandi R, et al. Spatially explicit effective reproduction numbers from incidence and mobility data. *Proceedings of the National Academy of Sciences*. 2023;120(20):e2219816120.
19. Abry P, Pustelnik N, Roux S, Jensen P, Flandrin P, Gribonval R, et al. Spatial and temporal regularization to estimate COVID-19 reproduction number $R(t)$: Promoting piecewise smoothness via convex optimization. *PLoS ONE*. 2020;15(8):e0237901.
20. Pascal B, Abry P, Pustelnik N, Roux S, Gribonval R, Flandrin P. Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data. *IEEE Transactions on Signal Processing*. 2022;70:2859–2868.

21. Pircalabelu E. A spline-based time-varying reproduction number for modelling epidemiological outbreaks. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 2023;72(3):688–702.
22. Ho F, Parag KV, Adam DC, Lau EH, Cowling BJ, Tsang TK. Accounting for the Potential of Overdispersion in Estimation of the Time-varying Reproduction Number. *Epidemiology*. 2023;34(2):201–205.
23. Azmon A, Faes C, Hens N. On the estimation of the reproduction number based on misreported epidemic data. *Statistics in Medicine*. 2014;33(7):1176–1192.
24. Gressani O, Faes C, Hens N. An approximate Bayesian approach for estimation of the instantaneous reproduction number under misreported epidemic data. *Biometrical Journal*. 2022;65(6):2200024.
25. Jin S, Dickens BL, Lim JT, Cook AR. EpiMix: A novel method to estimate effective reproduction number. *Infectious Disease Modelling*. 2023;8(3):704–716.
26. Hettinger G, Rubin D, Huang J. Estimating the instantaneous reproduction number with imperfect data: a method to account for case-reporting variation and serial interval uncertainty. arXiv preprint arXiv:230212078. 2023;.
27. Berry I, O'Neill M, Sturrock SL, Wright JE, Acharya K, Brankston G, et al. A sub-national real-time epidemiological and vaccination database for the COVID-19 pandemic in Canada. *Scientific Data*. 2021;8(1). doi:10.1038/s41597-021-00955-2.
28. Pitzer VE, Chitwood M, Havumaki J, Menzies NA, Perniciaro S, Warren JL, et al. The impact of changes in diagnostic testing practices on estimates of COVID-19 transmission in the United States. *American Journal of Epidemiology*. 2021;190(9):1908–1917.
29. Hitchings MD, Dean NE, García-Carreras B, Hladish TJ, Huang AT, Yang B, et al. The usefulness of the test-positive proportion of severe acute respiratory syndrome coronavirus 2 as a surveillance tool. *American Journal of Epidemiology*. 2021;190(7):1396–1405.

30. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LH, Fearon E, et al. Challenges in control of COVID-19: short doubling time and long delay to effect of interventions. *Philosophical Transactions of the Royal Society B*. 2021;376(1829):20200264.
31. Kim SJ, Koh K, Boyd S, Gorinevsky D. ℓ_1 trend filtering. *SIAM Review*. 2009;51(2):339–360.
32. Tibshirani RJ. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*. 2014;42(1):285–323.
33. Tibshirani RJ. Divided differences, falling factorials, and discrete splines: Another look at trend filtering and related problems. *Foundations and Trends® in Machine Learning*. 2022;15(6):694–846.
34. Sadhanala V, Bassett R, Sharpnack J, McDonald DJ. Exponential Family Trend Filtering on Lattices. *arXiv preprint arXiv:220909175*. 2022;;
35. Vaiter S, Deledalle C, Fadili J, Peyré G, Dossal C. The degrees of freedom of partly smooth regularizers. *Annals of the Institute of Statistical Mathematics*. 2017;69:791–832.
36. Groendyke C, Welch D, Hunter DR. Bayesian inference for contact networks given epidemic data. *Scandinavian Journal of Statistics*. 2011;38(3):600–616.
37. Cori A, Boëlle PY, Thomas G, Leung GM, Valleron AJ. Temporal variability and social heterogeneity in disease transmission: the case of SARS in Hong Kong. *PLoS computational biology*. 2009;5(8):e1000471.
38. BCCDC. BC Centre for Disease Control Covid-19 Dashboard Case; 2023.
39. Lehtinen S, Ashcroft P, Bonhoeffer S. On the relationship between serial interval, infectiousness profile and generation time. *Journal of the Royal Society Interface*. 2021;18(174):20200756.
40. Frost WH, Sydenstricker E. Influenza in Maryland: preliminary statistics of certain localities. *Public Health Reports (1896-1970)*. 1919; p. 491–504.

41. Taubenberger JK, Morens DM. 1918 Influenza: the mother of all pandemics. Emerging Infectious Diseases. 2006;17(1):69–79.

Supplementary details on experiments of effective reproduction number estimation with trend filtering

Jiapeng Liu, Zhenglun Cai, Paul Gustafson, and Daniel J. McDonald

A.1 Derivation of Kullback Leibler divergence for accuracy comparison

A.2 Supplmentary details on experimental settings

We compare the accuracy of the estimated effective reproduction numbers using Kullback Leibler (KL) divergence (with Poisson distributional assumption on incidence) averaged per coordinate across our `RtEstim` and several alternative methods including `EpiEstim` with weekly and monthly sliding windows, `EpiLPS`, `EpiFilter`, `EpiNow2`, and `RtEstim` with degrees $k=0,1,2,3$, which yields 9 methods in total. We consider two lengths of epidemics with $n = 50, 300$ timepoints respectively. Since `EpiNow2` runs too long (specifically, for one long epidemic, it takes ???TBA), we only compare it with other methods for short epidemics.

We consider serial interval (SI) distributions of measles and SARS to generate long synthetic epidemics, and flu for short epidemics. Incident cases in synthetic measles epidemics are relatively low (within 1000 at the peak overall), and SARS incident cases are relatively large (between 15000 and 20000 at the peak overall). We specifically consider a reasonably large overdispersion level of negative Binomial incidence is of size 5. Figure A.2.1 displays the ratio of standard deviation over mean (called, sigma to mean ratio) of incidence across different settings using the same set of sample epidemics in the manuscript. Compared to the counterpart of Poisson incidence (which decreases quickly to be under 0.25 and close to 0 at most timepoints) per \mathcal{R}_t scenario for each epidemic, the negative Binomial incidence appears to have an apparently larger sigma to mean ratio (staying at around 0.5 at most timepoints) implying a distinguishable overdispersion level.

In model fitting, we use both true and misspecified serial interval (SI) distributions. The misspecification of serial interval distributions are either mild or major, where, in the major misspecification, we use SI of SARS to solve long measles epidemics and SI of measles to solve short flu epidemics. While, in the mild SI misspecification, we consider a shaped (mean increased by 2) and scaled (standard deviation increased by 10) parameters for both short flu and long measles epidemics, denoted as `flu_ss` and `measles_ss` respectively. These settings result in 7 pairs of SI distributions (for epidemic generating, and for model fitting), i.e., (`measles, measles`), (`SARS, SARS`), (`measles, measles_ss`), (`measles, SARS`) for long epidemics and (`flu, flu`), (`flu, flu_ss`), (`flu, measles`) for short epidemics. Figure A.2.2 displays the SI distributions (`measles, measles_ss, SARS, flu, and flu_ss`) used in the experiments.

Table 1 summarizes the aforementioned experimental setting for accuracy comparison. Poisson and negative Binomial (NB) distributions for incidence and 4 \mathcal{R}_t scenarios are used for all settings. Each experimental

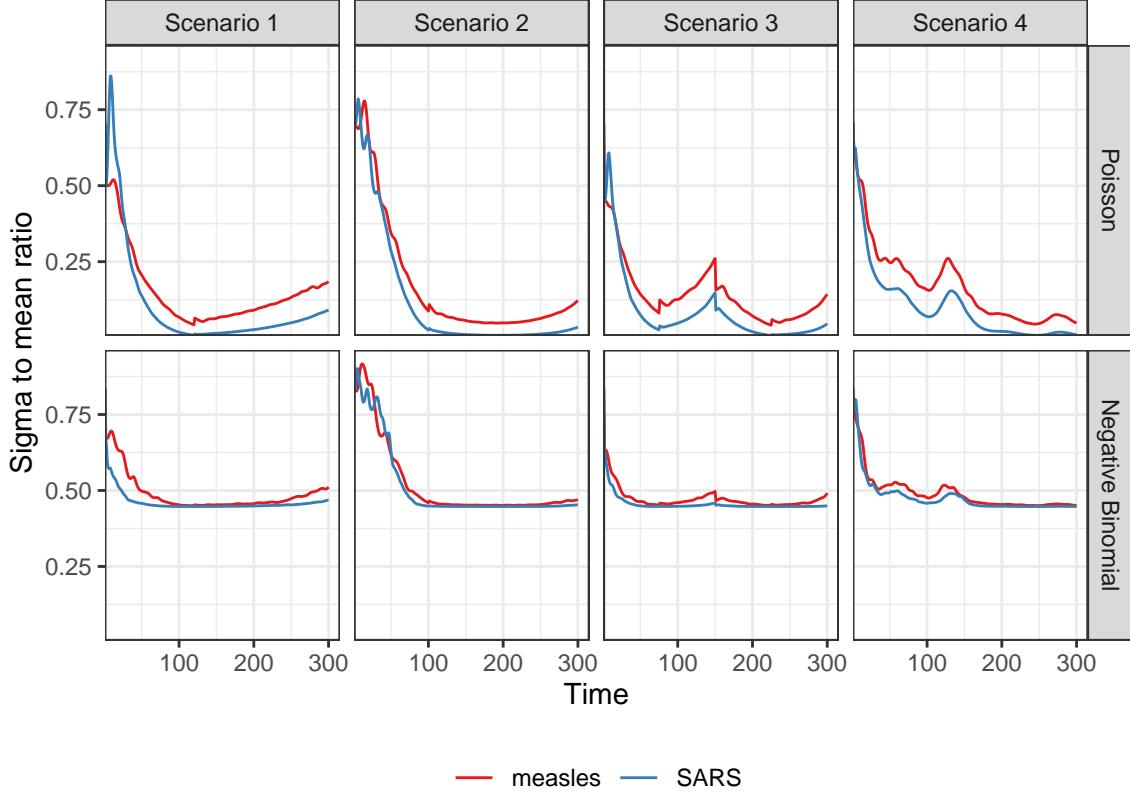


Figure A.2.1: Dispersion level of incidence of sample epidemics

Table 1: Summary of experimental setting on accuracy comparison

Length	SI	Rt scenario	Incidence	SI for modelling	Method
300	measles	1-4	Poisson, NB	measles, measles_ss, SARS	8 methods
300	SARS	1-4	Poisson, NB	SARS	8 methods
50	flu	1-4	Poisson, NB	flu, flu_ss, measles	9 methods

setting is replicated for 50 times, which yields 12800 experiments with long epidemics and 2700 with short epidemics.

Here we list the hyperparameters used in modelling for each method. Most of them are the experimental settings used in the papers where they were proposed and deemed as the “best” tuned ones. We consider both weekly and monthly sliding windows in `EpiEstim`, 40 basis functions in `EpiLPS` with the NelderMead method to maximize the hyperparameter posterior distribution. We input 2000 grid size in `EpiFilter` with 0.1 diffusion noise and uniform prior on \mathcal{R}_t with mean 1/2000, and use the smoothed \mathcal{R}_t as estimates. We run 10-fold cross validation (CV) to choose the best tuning parameter from the candidate set of size 50, i.e., $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_{50}\}$. Specifically, we divide all samples (except the first and last entries) into ten folds evenly and randomly, and build models on each sample set by leaving a fold out across all hyperparameters. We select the tuning parameter that gives the lowest averaged **deviance** between the estimated incidence and the observed samples averaged over all folds.

We visualize the selected key results of accuracy comparison using long synthetic epidemics in Section 3.2 in the manuscript. Other main experimental results are displayed as follows.

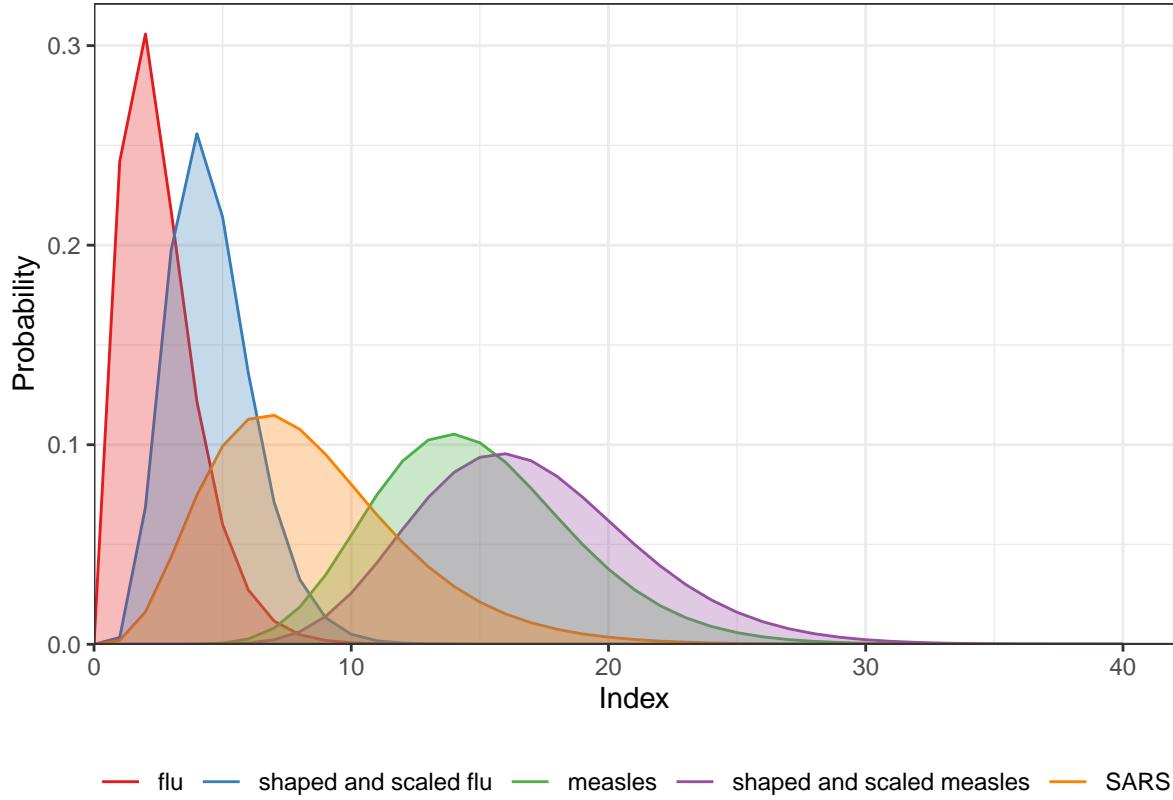


Figure A.2.2: Density curves of serial interval distributions used in the experiments

A.3 Supplementary experimental results on accuracy comparison

Figures A.3.1 and A.3.2 display the KL divergence excluding the first months for measles and SARS epidemics respectively. To compare **EpiEstim** with *monthly* sliding windows with other methods, we average the KL divergence per coordinate excluding the timepoints in the first months for all approaches, since **EpiEstim** estimates with the monthly sliding windows are not available until the second months. The *y*-axis is displayed on a logarithmic scale for a better visualization, since a few values are much larger than others.

The relative performance of **EpiEstim** with monthly sliding windows, in general, is not as good as its weekly sliding window based on the relative positions of its boxes and the counterparts of the other methods, except for the Scenario 2 with negative Binomial incidence. It can be explained by **EpiEstim** with longer sliding windows assume similarity of neighbouring \mathcal{R}_t across longer periods, and thus, is smoother and less accurate compared to the one with shorter sliding windows.

Figures A.3.3 and A.3.4 display the KL divergence for short epidemics averaged per coordinate excluding the first weeks and months respectively.

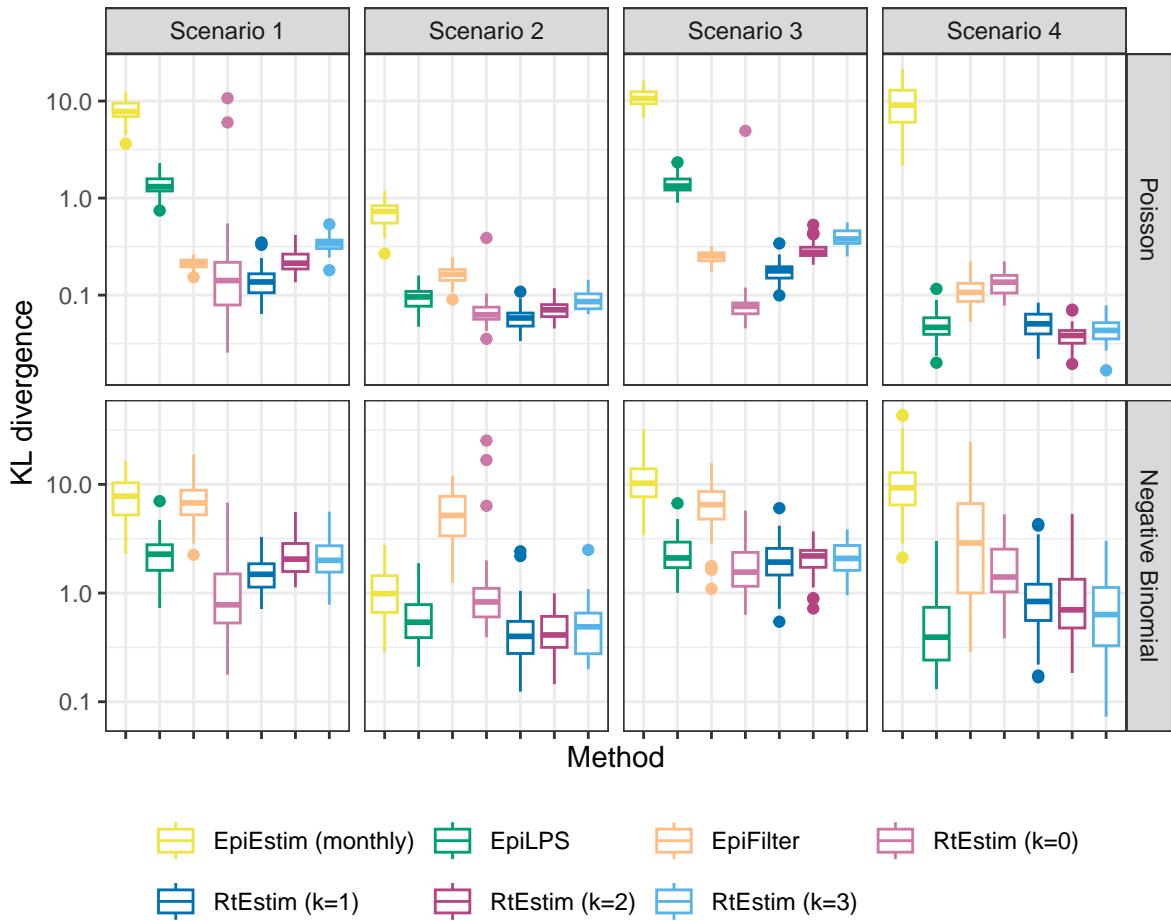


Figure A.3.1: KL divergence excluding the first months for measles epidemics. Y-axis is on a logarithmic scale.

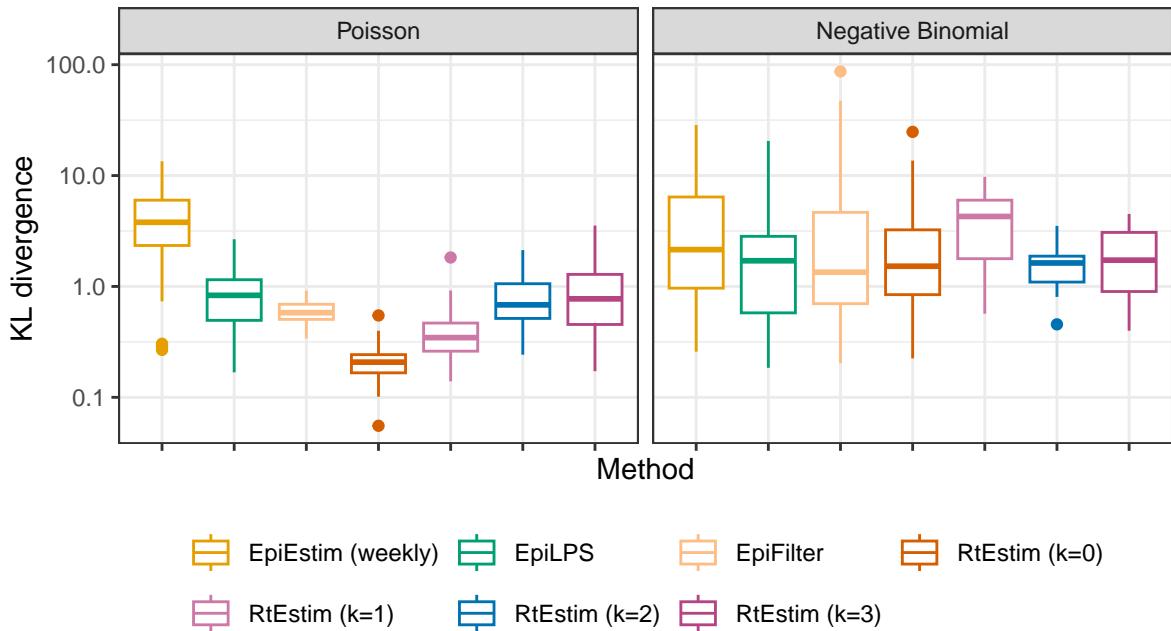


Figure A.3.3: KL divergence excluding the first weeks for flu epidemics. Y-axis is on a logarithmic scale.

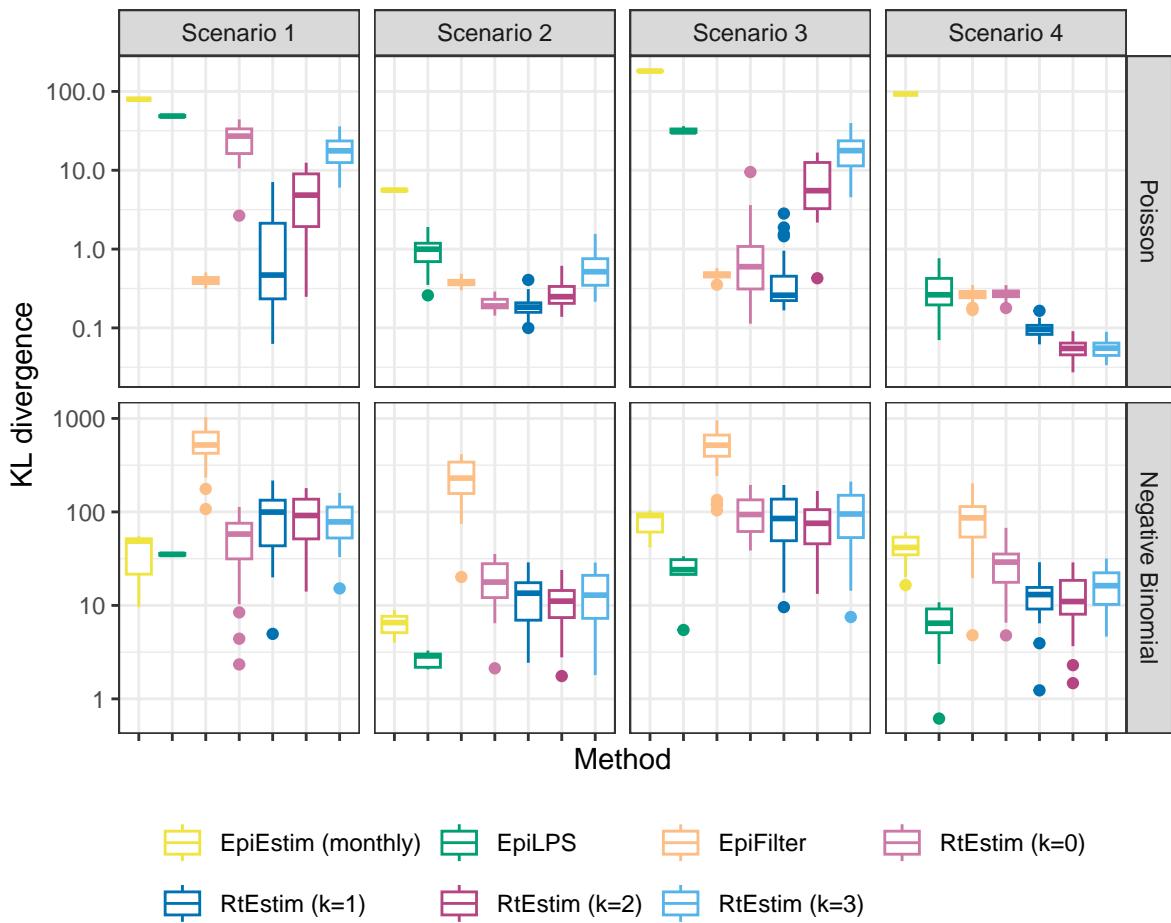


Figure A.3.2: KL divergence excluding the first months for SARS epidemics. Y-axis is on a logarithmic scale.

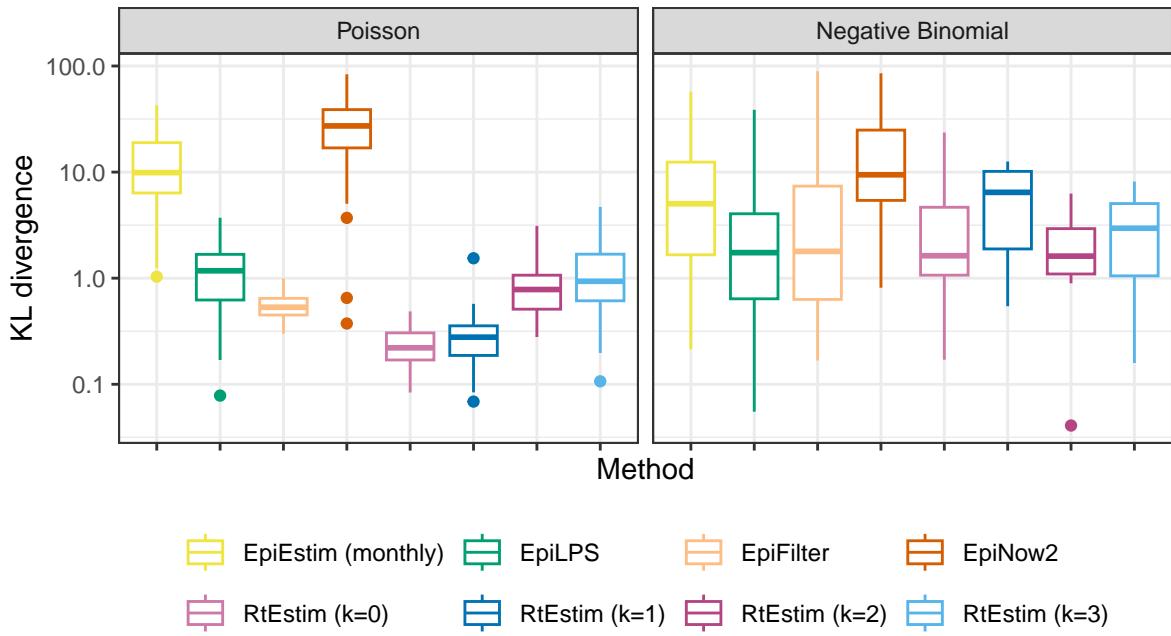


Figure A.3.4: KL divergence excluding the first months for flu epidemics. Y-axis is on a logarithmic scale.

A.4 Experimental results on accuracy under misspecification of serial interval distributions

A.4.1 Mild misspecification

Figures x and x display KL divergence (excluding first weeks and months respectively) for all 8 methods with shaped and scaled `measles` SI parameters with long `measles` epidemics across all settings.

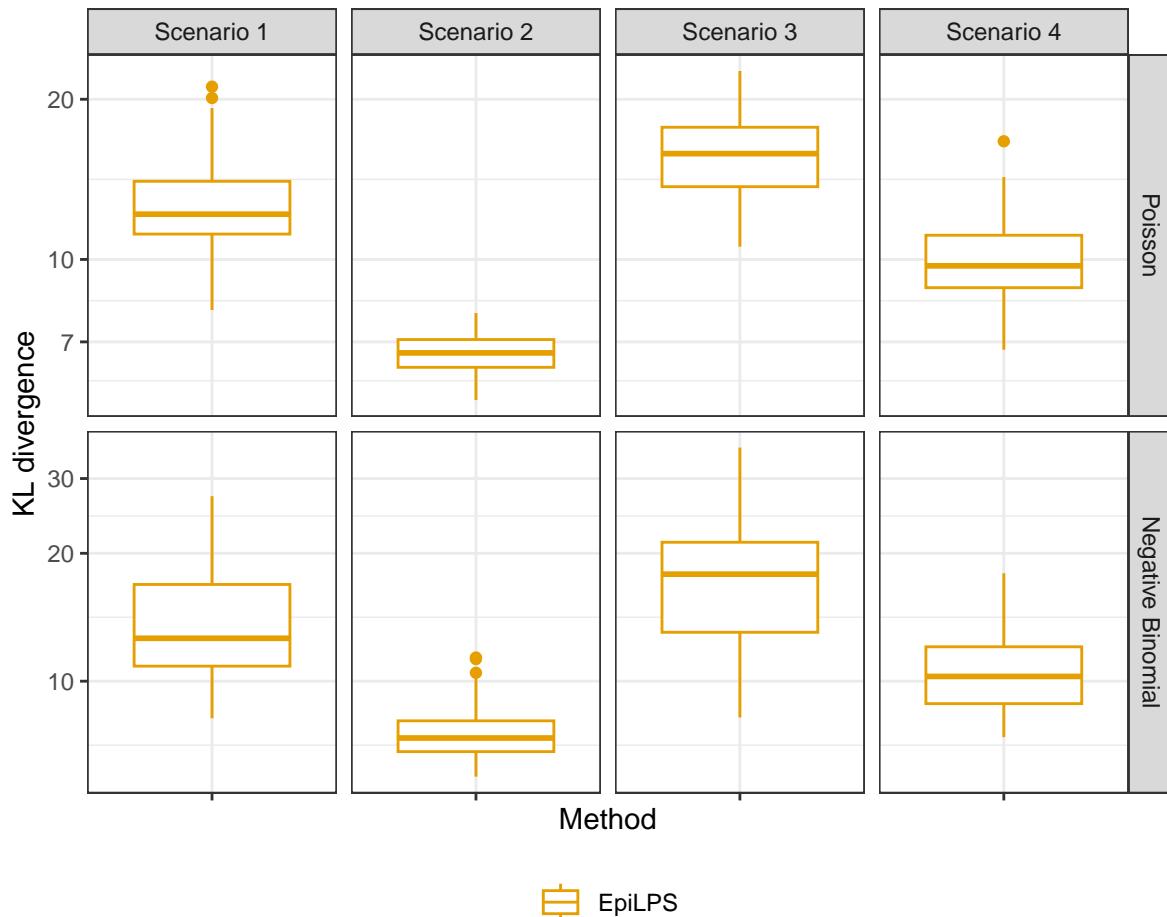


Figure A.4.1: KL divergence excluding the first weeks for measles epidemics with mild SI misspecification. Y-axis is on a logarithmic scale.

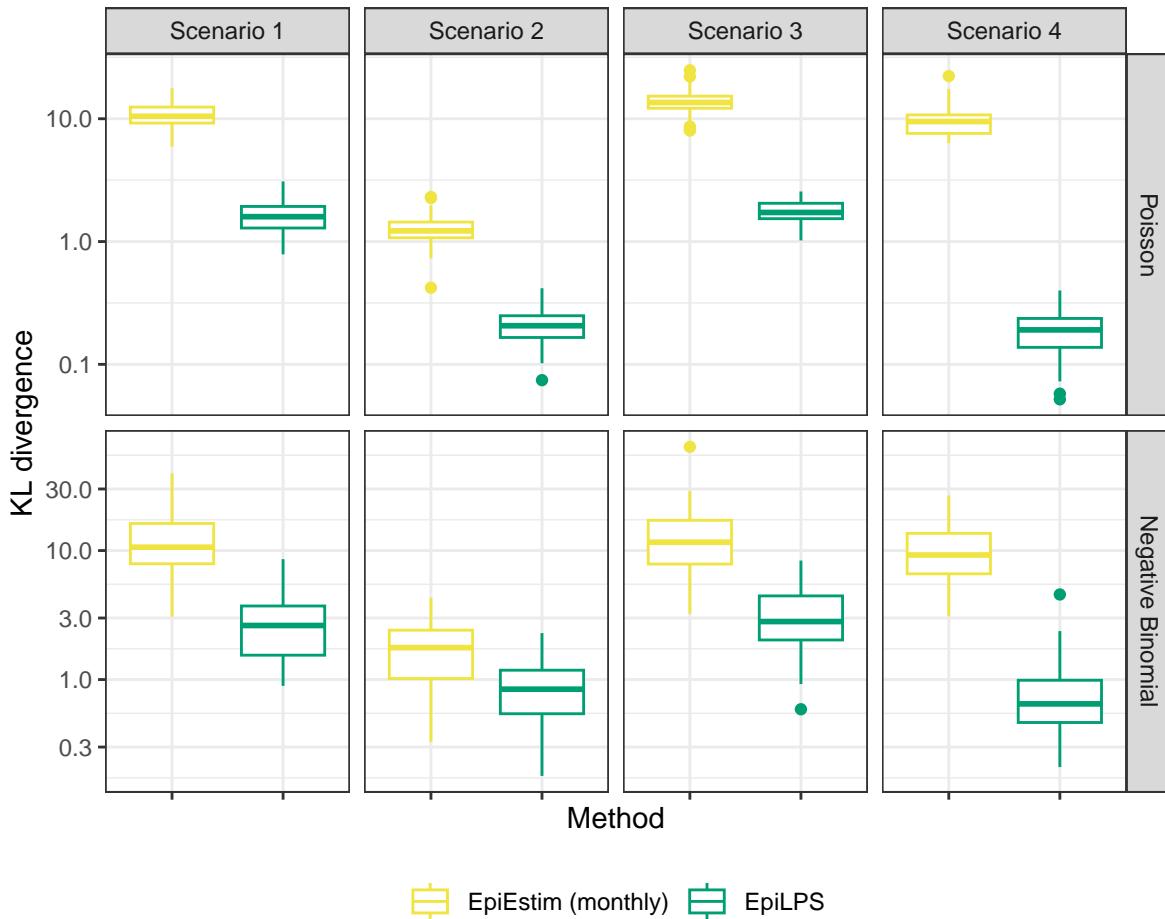


Figure A.4.2: KL divergence excluding the first months for measles epidemics with mild SI misspecification. Y-axis is on a logarithmic scale.

Figures x and x display KL divergence (excluding first weeks and months respectively) for all 9 methods with shaped and scaled `flu` SI parameters with short `flu` epidemics across all settings.

A.4.2 Major misspecification

Figures A.4.3 and A.4.4 display KL divergence (excluding first weeks and months respectively) for all 8 methods with `SARS` SI parameters with long `measles` epidemics across all settings.

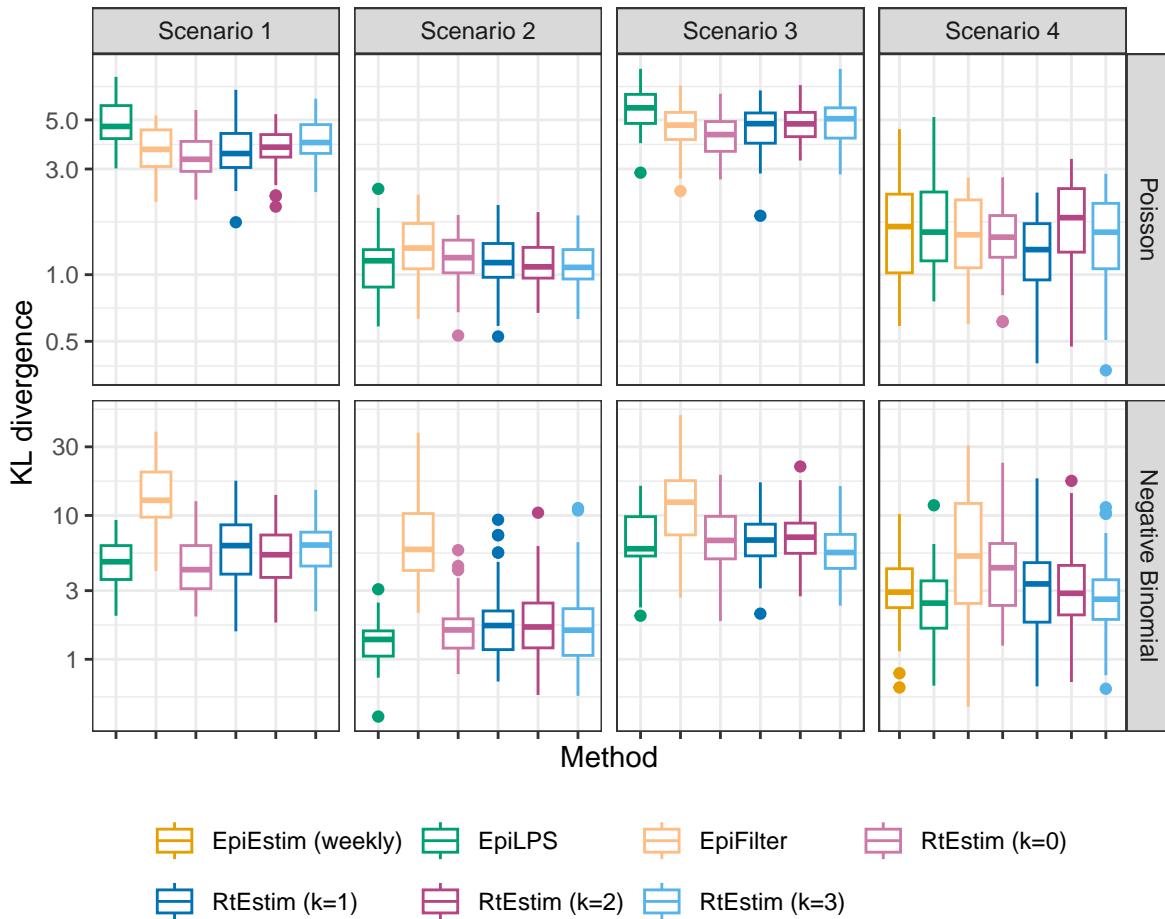


Figure A.4.3: KL divergence excluding the first weeks for measles epidemics with major SI misspecification. Y-axis is on a logarithmic scale.

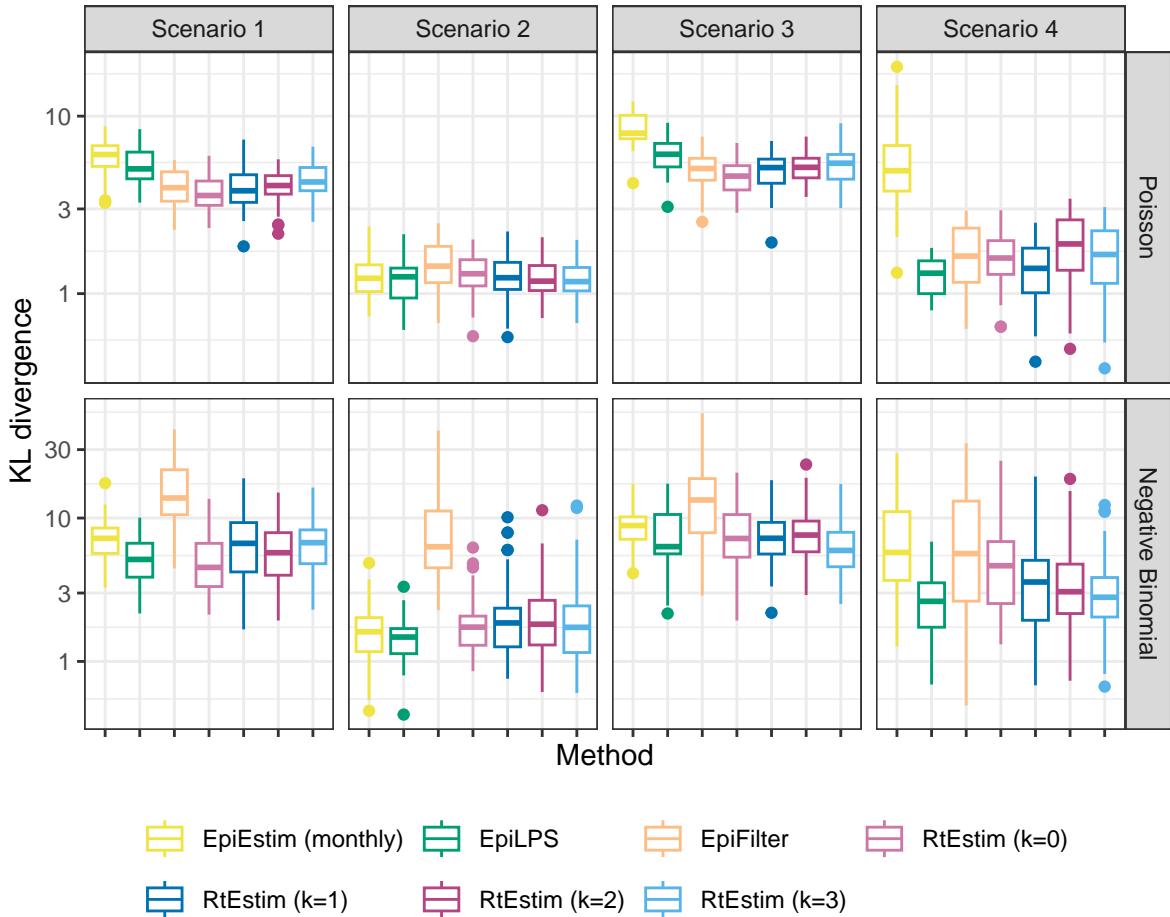


Figure A.4.4: KL divergence excluding the first months for measles epidemics with major SI misspecification. Y-axis is on a logarithmic scale.

Figures A.4.5 and A.4.6 display KL divergence (excluding first weeks and months respectively) for all 9 methods with `measles` SI parameters with short `flu` epidemics across all settings.

A.5 Confidence interval coverage

A.5.1 Display estimates and confidence intervals for sample epidemics

Let's review the estimated \mathcal{R}_t with 95% confidence intervals for the sample long epidemics by all methods in the manuscript in Figure xx.

\begin{figure}[H]

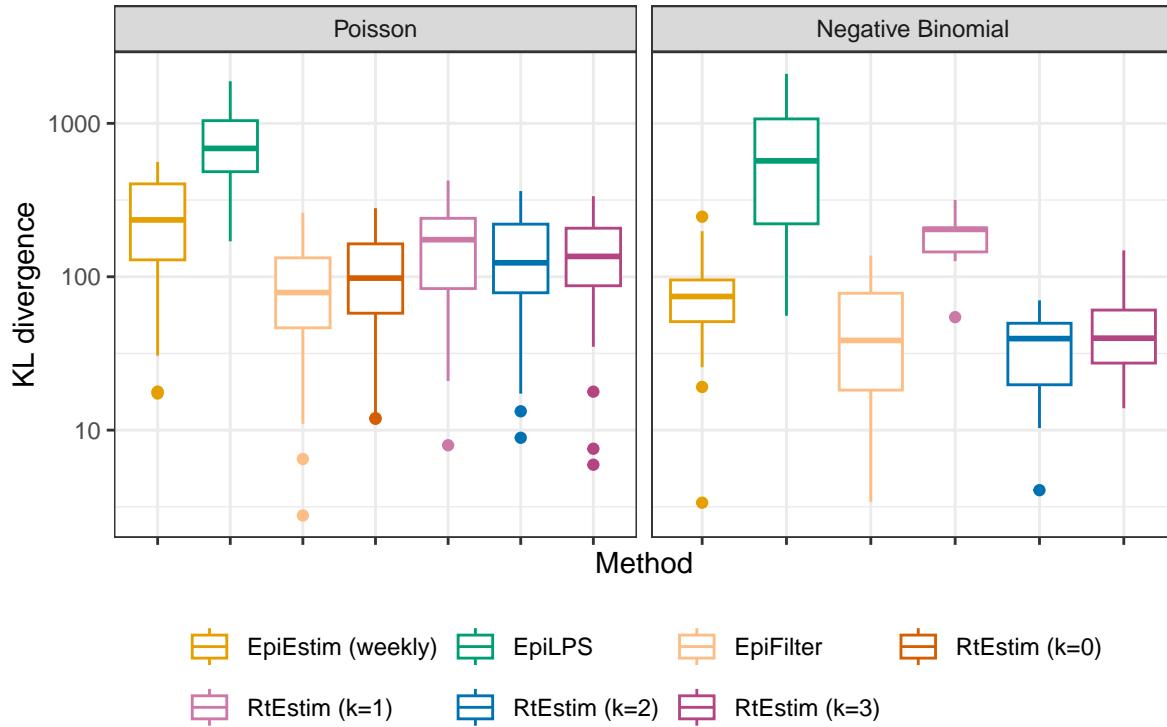


Figure A.4.5: KL divergence excluding the first weeks for flu epidemics with major SI misspecification. Y-axis is on a logarithmic scale.

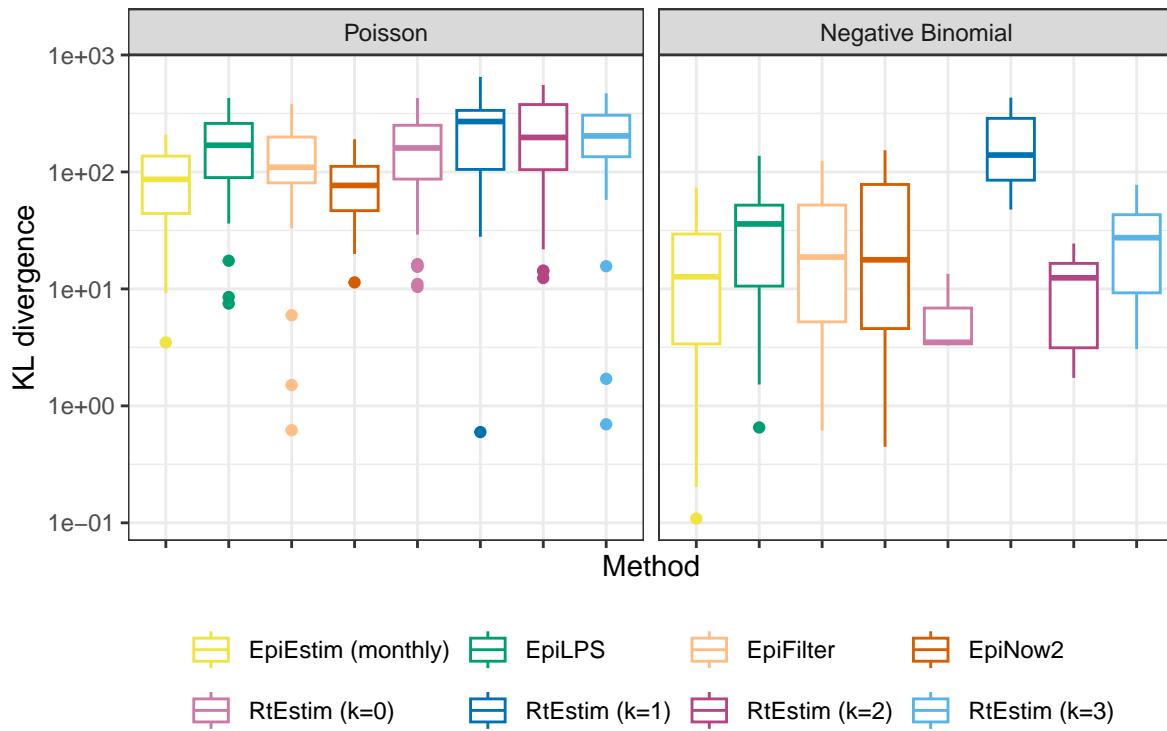
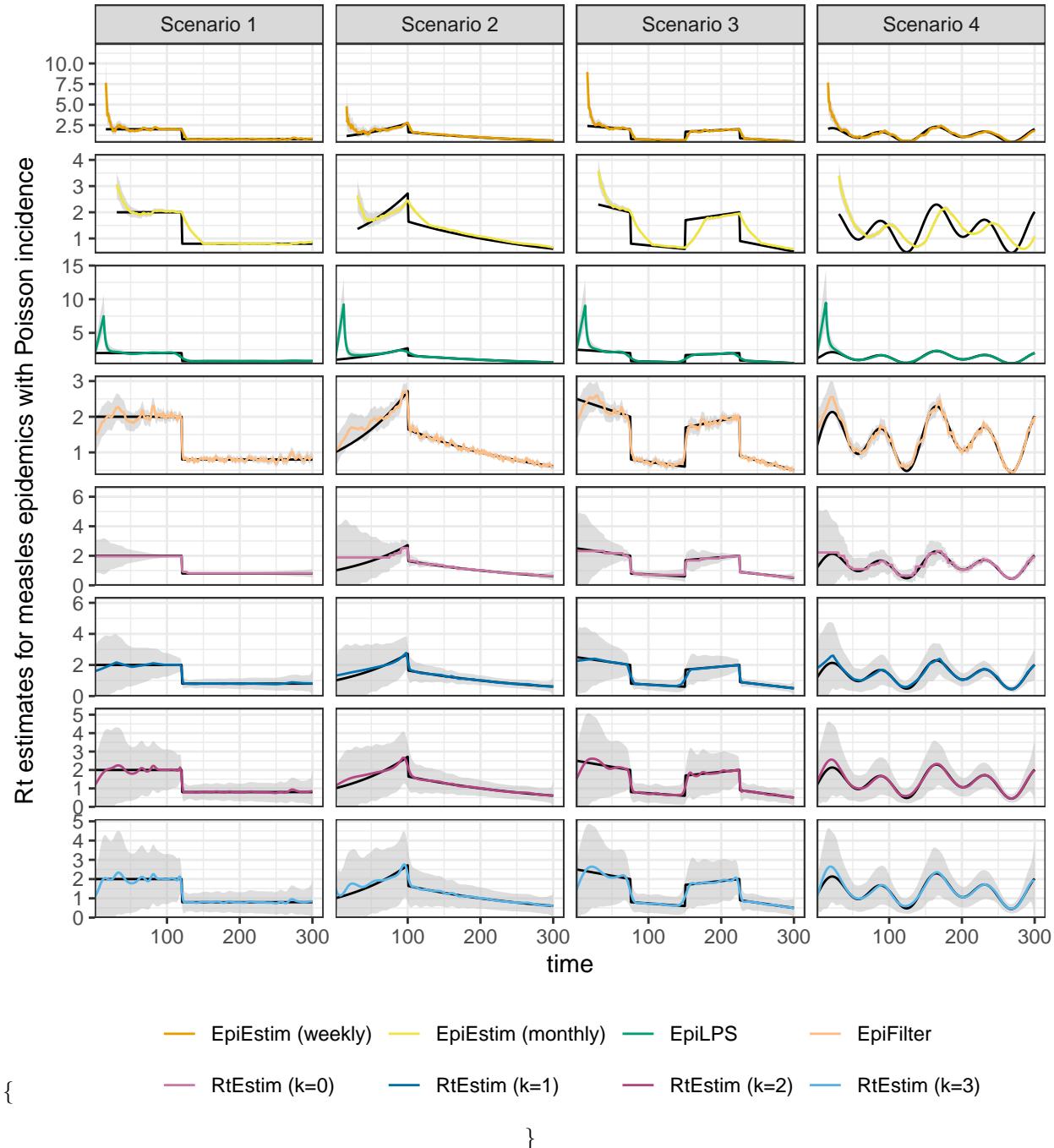


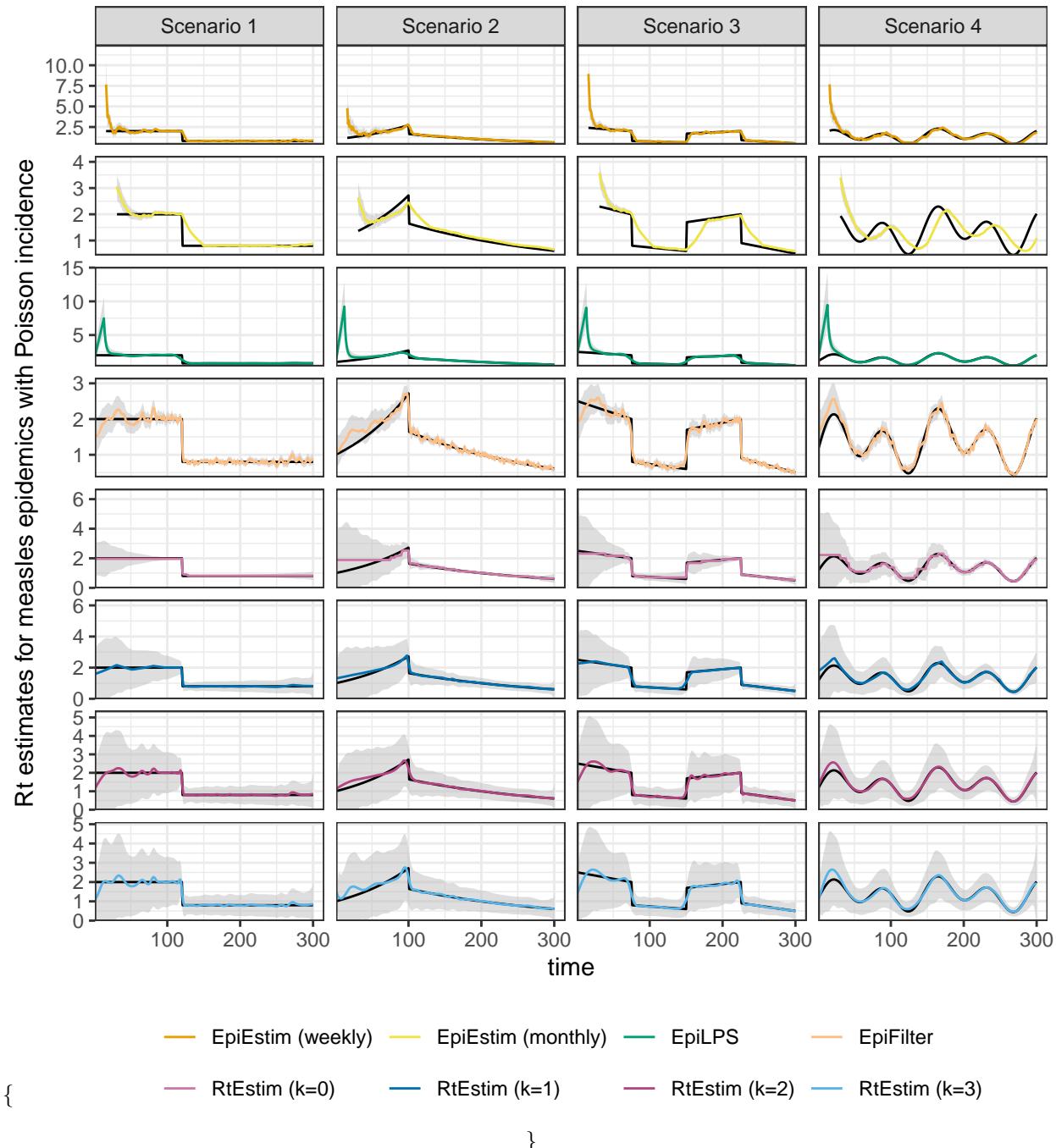
Figure A.4.6: KL divergence excluding the first months for flu epidemics with major SI misspecification. Y-axis is on a logarithmic scale.

A.5.1 Display estimates and confidence intervals for sample epidemic CONFIDENCE INTERVAL COVERAGE



\caption{Fitted Rt with 95% confidence intervals for measles epidemics with Poisson incidence.} \end{figure}

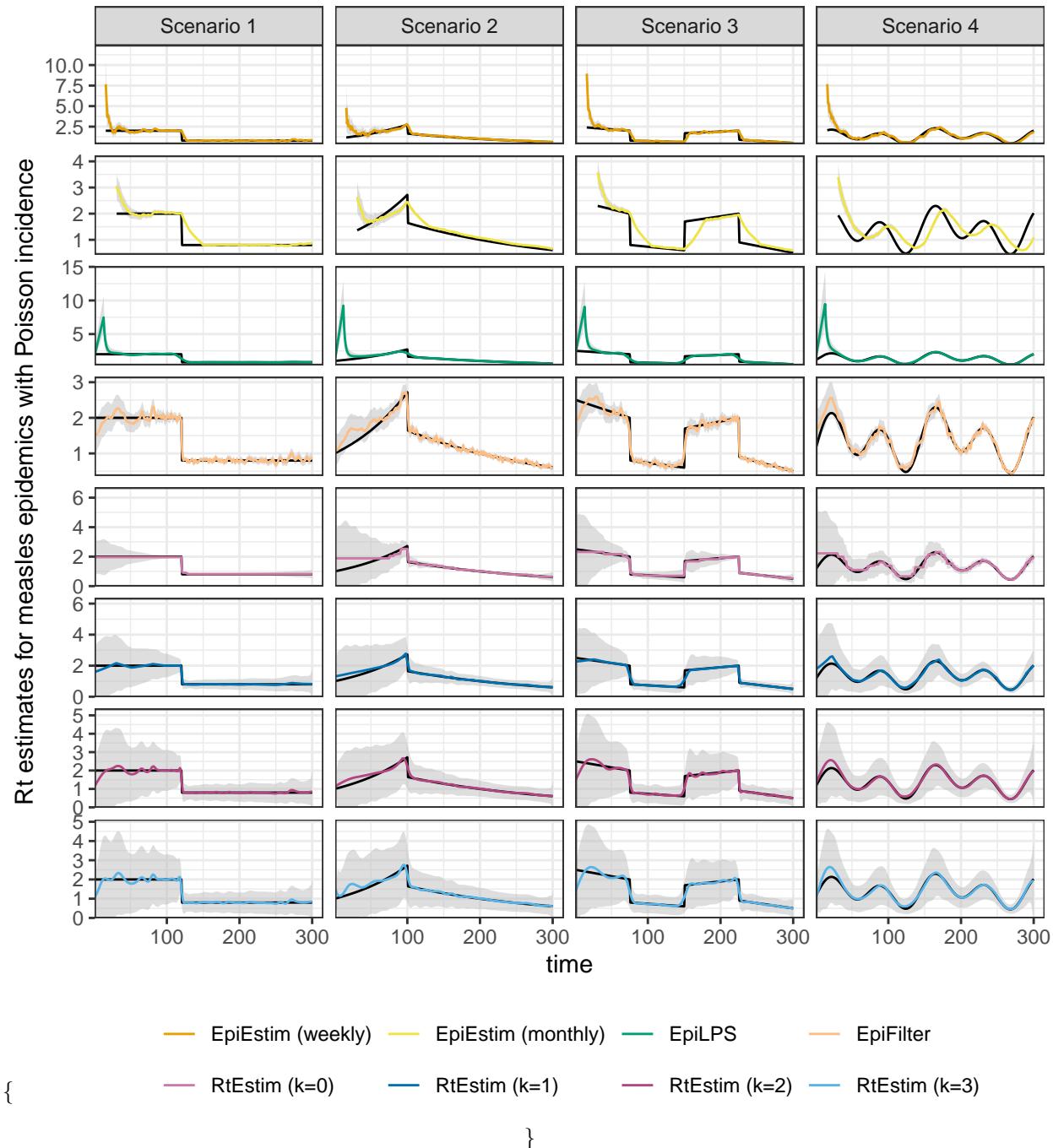
\begin{figure}[H]



\caption{Fitted Rt with 95\% confidence intervals for measles epidemics with negative Binomial incidence.}

\end{figure}

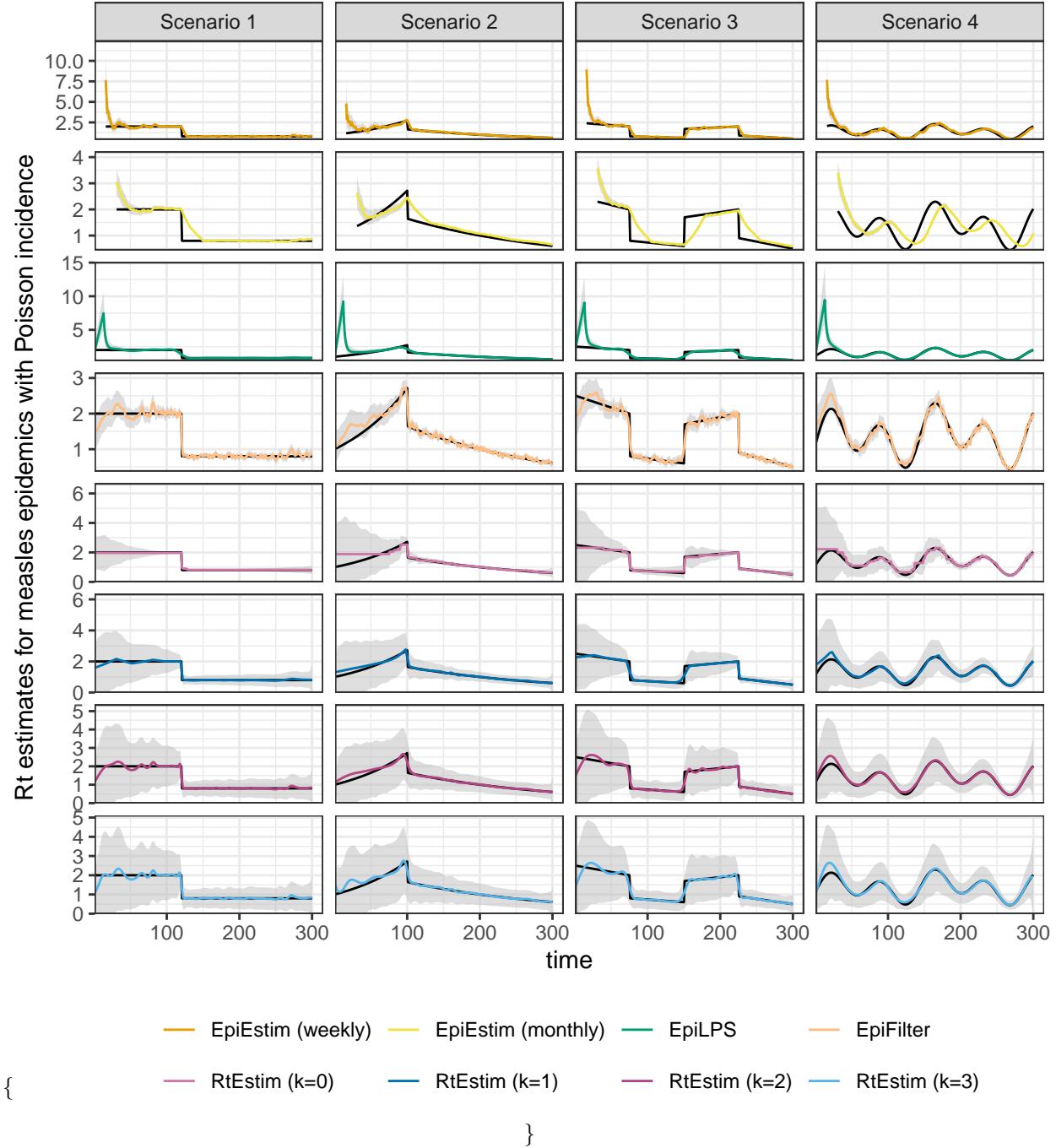
\begin{figure}[H]



\caption{Fitted Rt with 95% confidence intervals for SARS epidemics with negative Binomial incidence.}

\end{figure}

\begin{figure}[H]



\caption{Fitted Rt with 95% confidence intervals for SARS epidemics with negative Binomial incidence.}
\end{figure}

A.5.2 Experimental settings on coverage level comparison of confidence intervals

We focus on a specific \mathcal{R}_t scenario, the piecewise linear case, and only long epidemics to compare the coverage of 95% confidence intervals across all 8 methods. We use the true serial interval distributions in modelling. Table 2 summarizes the experimental settings.

Table 2: Summary of experimental setting on coverage of confidence intervals

Length	SI	Rt scenario	Incidence	SI for modelling	Method
300	measles	3	Poisson, NB	measles	8 methods
300	SARS	3	Poisson, NB	SARS	8 methods

A.5.3 Experimental results

Figures A.5.1 and A.5.2 displays the percentages of coverage of 95% CI per coordinate over 50 random samples for measles and SARS epidemics respectively.

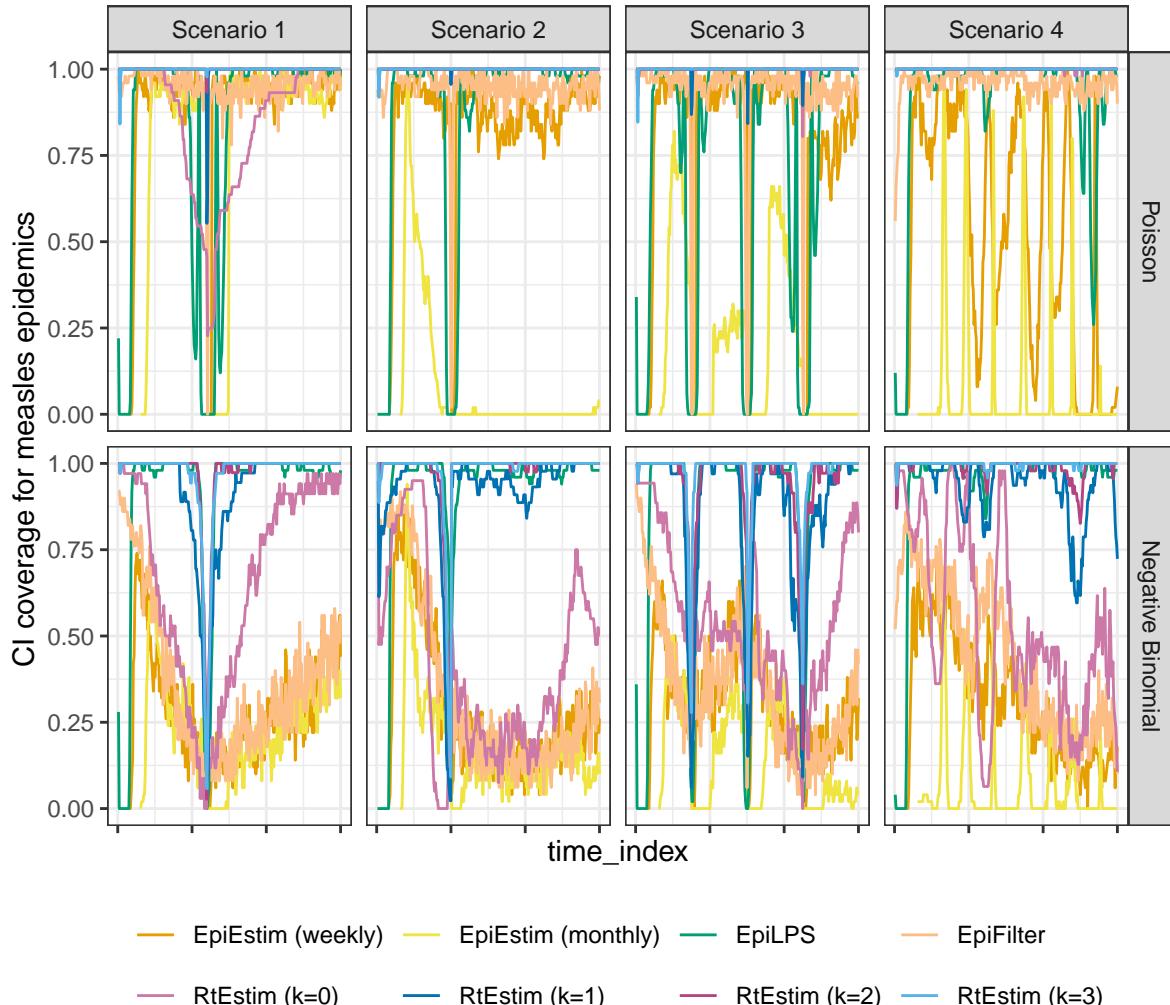


Figure A.5.1: Averaged coverage of CI per coordinate with measles epidemics.

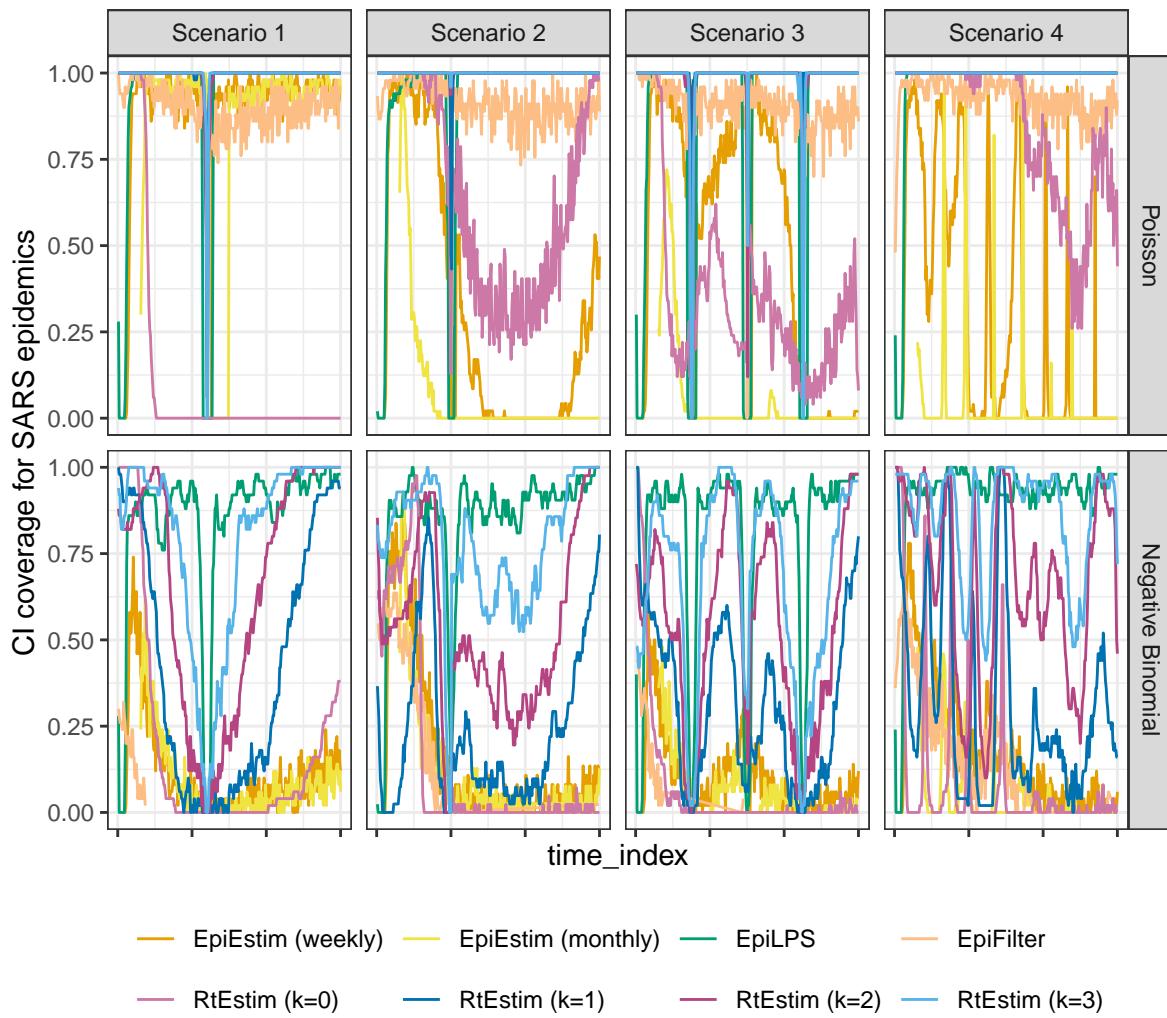
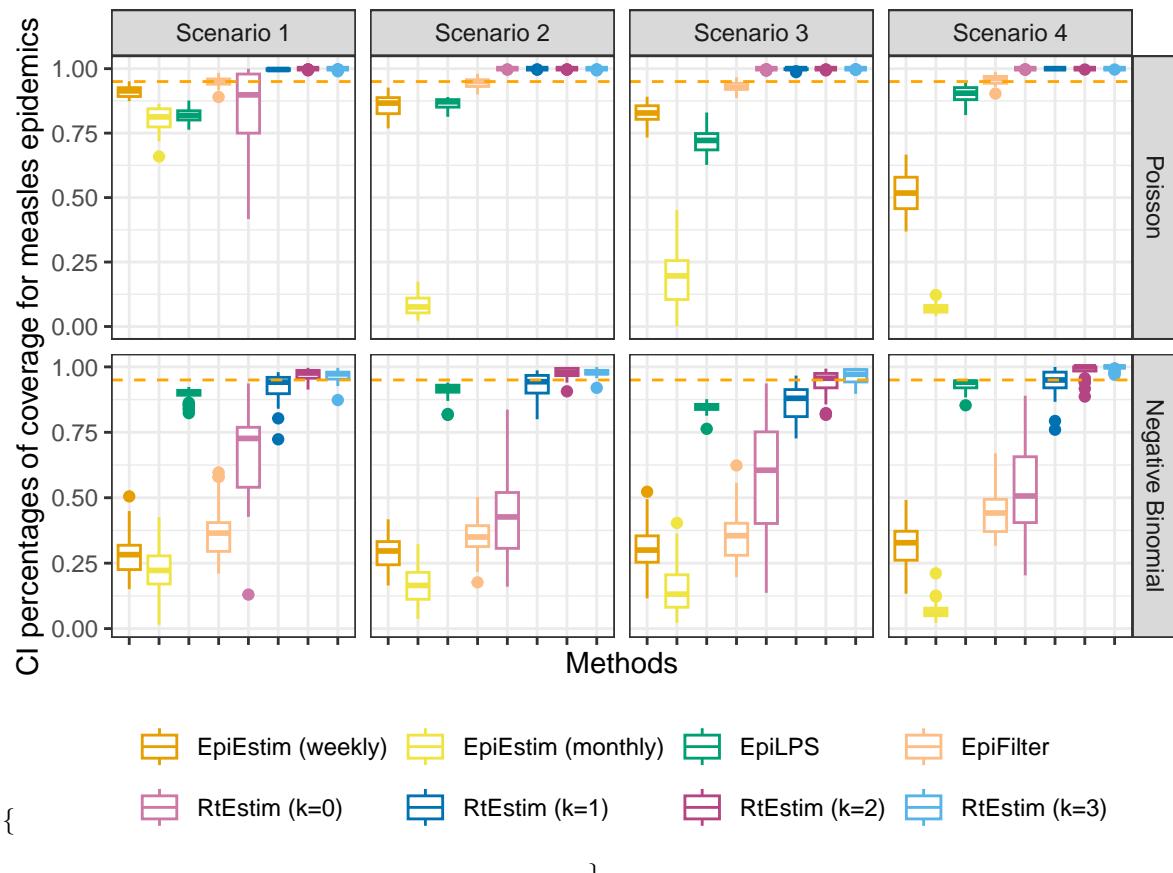


Figure A.5.2: Averaged coverage of CI per coordinate with SARS epidemics.

Figures A.5.3 and A.5.3 displays the percentages of coverage of 95% CI across all timepoints averaged over 50 random measles and SARS epidemics respectively.

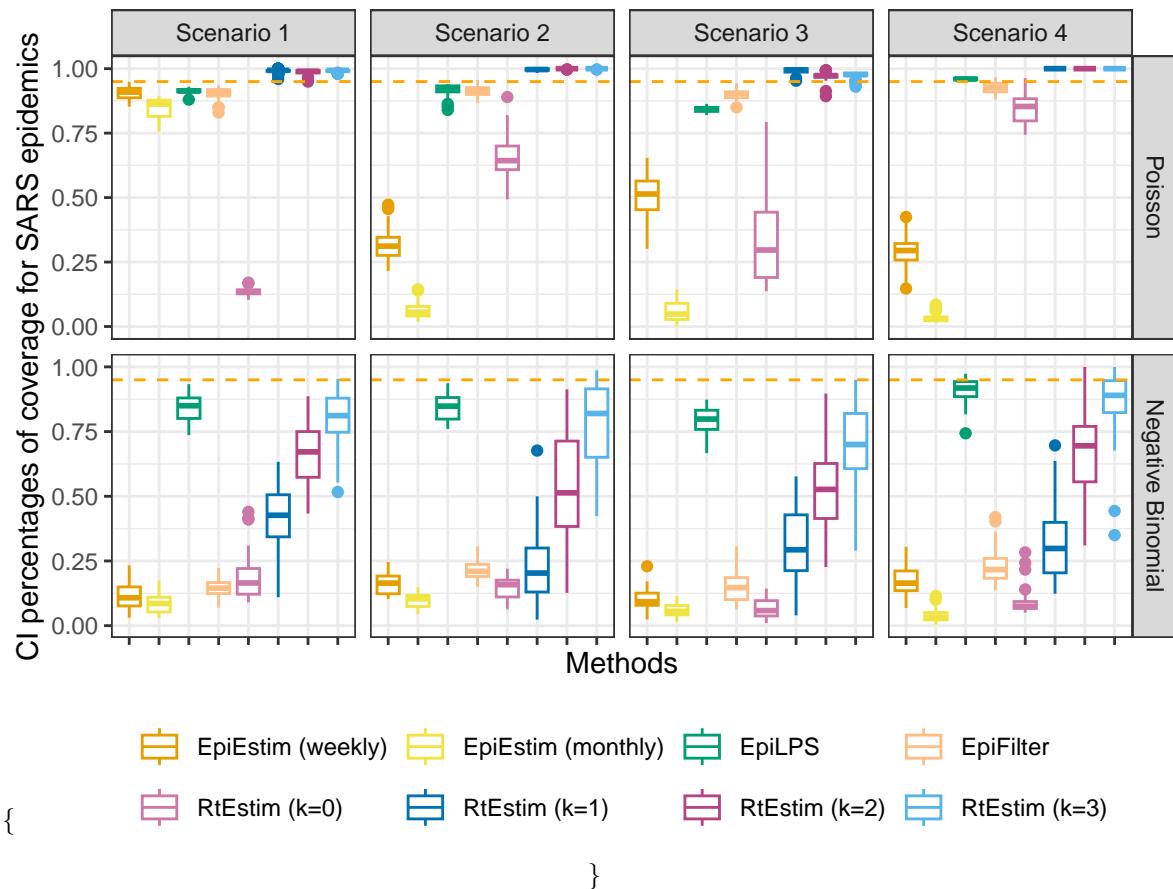
\begin{figure}[H]



\caption{Averaged percentages of 95% CI with measles epidemics.} \end{figure}

"

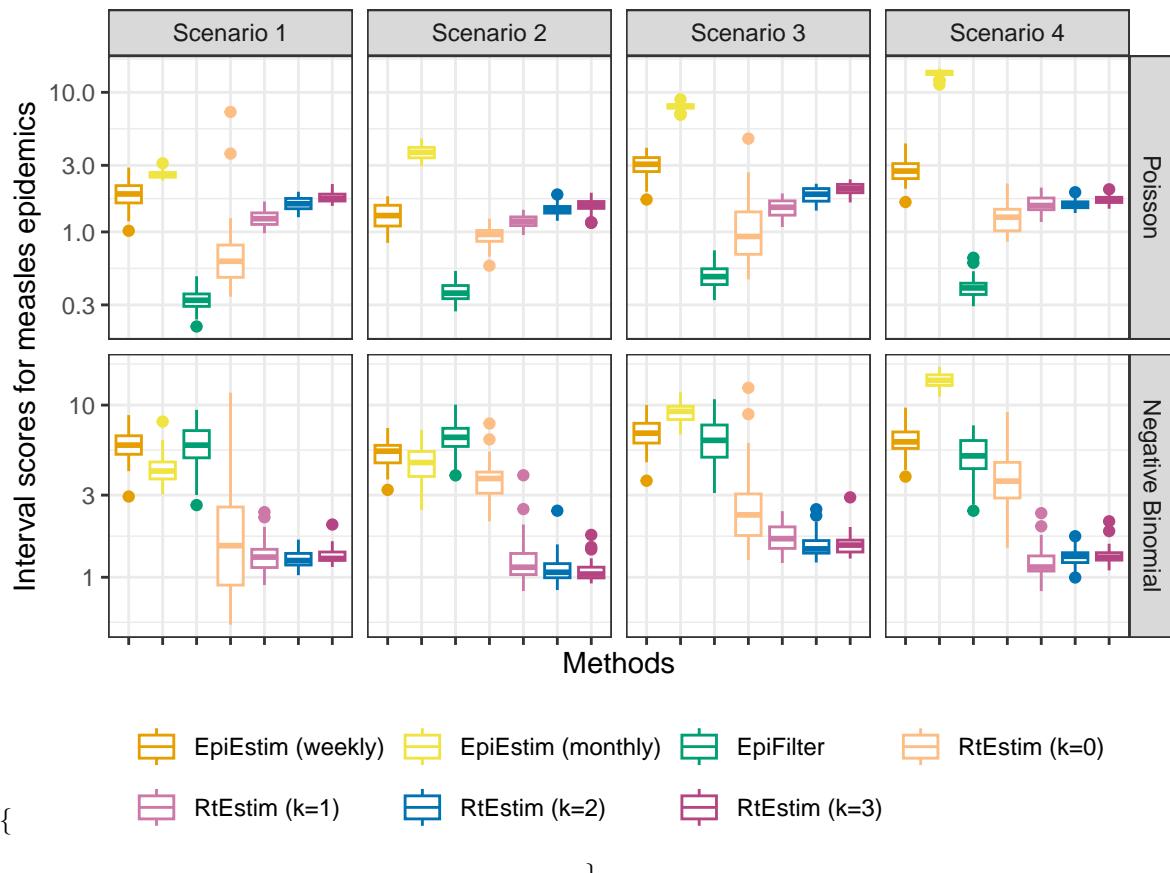
\begin{figure}[H]



\caption{Averaged percentages of 95% CI with SARS epidemics.} \end{figure}

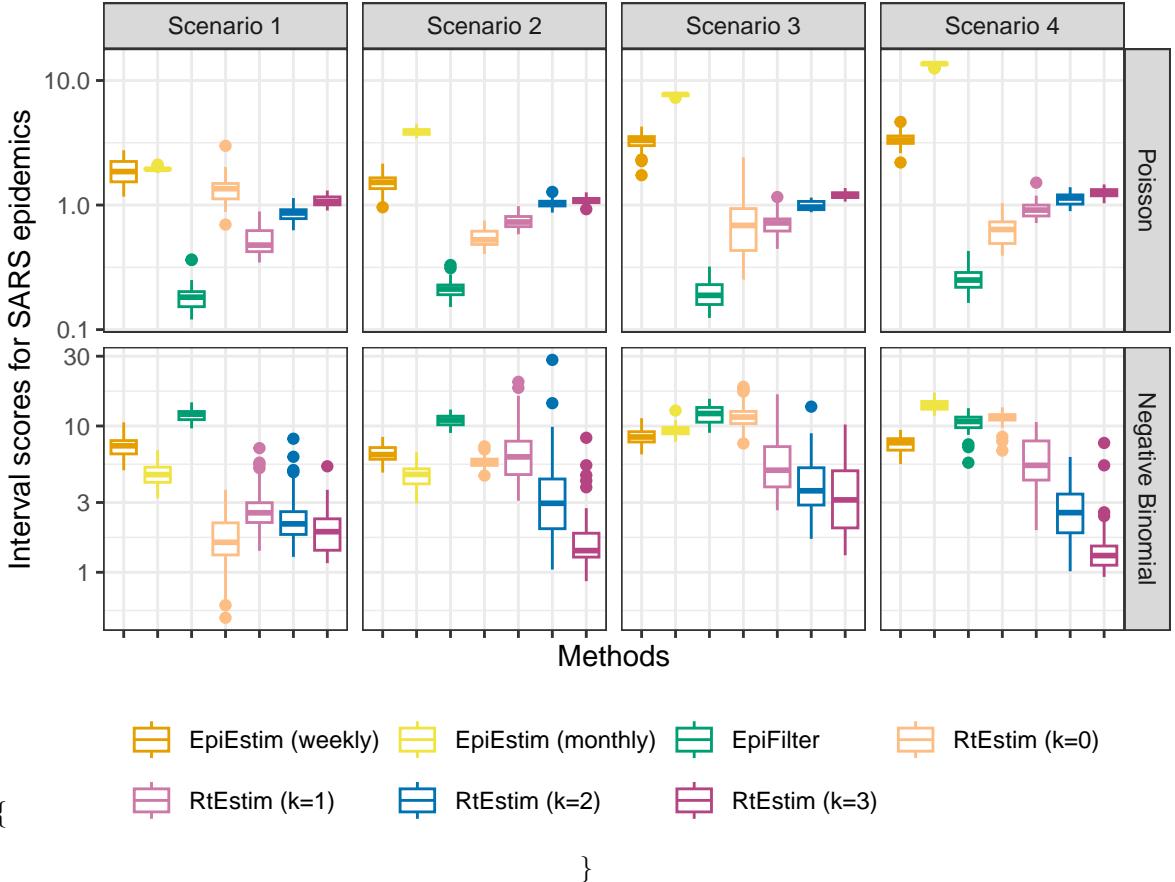
Figures A.5.3 and A.5.3 displays the interval scores of 95% CI averaged over 50 random measles and SARS epidemics respectively.

\begin{figure}[H]



\caption{Averaged interval scores of 95% CI with measles epidemics.}

\begin{figure}[H]



\caption{Averaged interval scores of 95\% CI with SARS epidemics.} \end{figure}

A.6 Time comparisons of methods for Section 3.2

Figures A.6.1 and A.6.2 show the time comparisons across all methods. `EpiEstim` with both weekly and monthly sliding windows are very fast and converge in less than 0.1 seconds. Piecewise constant `RtEstim` (with $k=0$) estimates can be generated within 0.1 seconds as well. `EpiLPS` is slightly slower, but still very fast and within 1 second for all experiments. Piecewise linear and cubic `RtEstim` (with $k = 1$ and $k = 3$ respectively) are slower, but mostly within 10 seconds.

It is remarkable that our `RtEstim` computes 50 lambda values with 10-fold CV for each experiment, which results in 550 times of modelling per experiment (including modelling for all folds). The running times are no more than 10 seconds for most of the experiments, which means the running time for each time of modelling is very fast, and on average can be less than 0.02 seconds. The other two methods only run once for a fixed set of hyperparameters for each experiment.

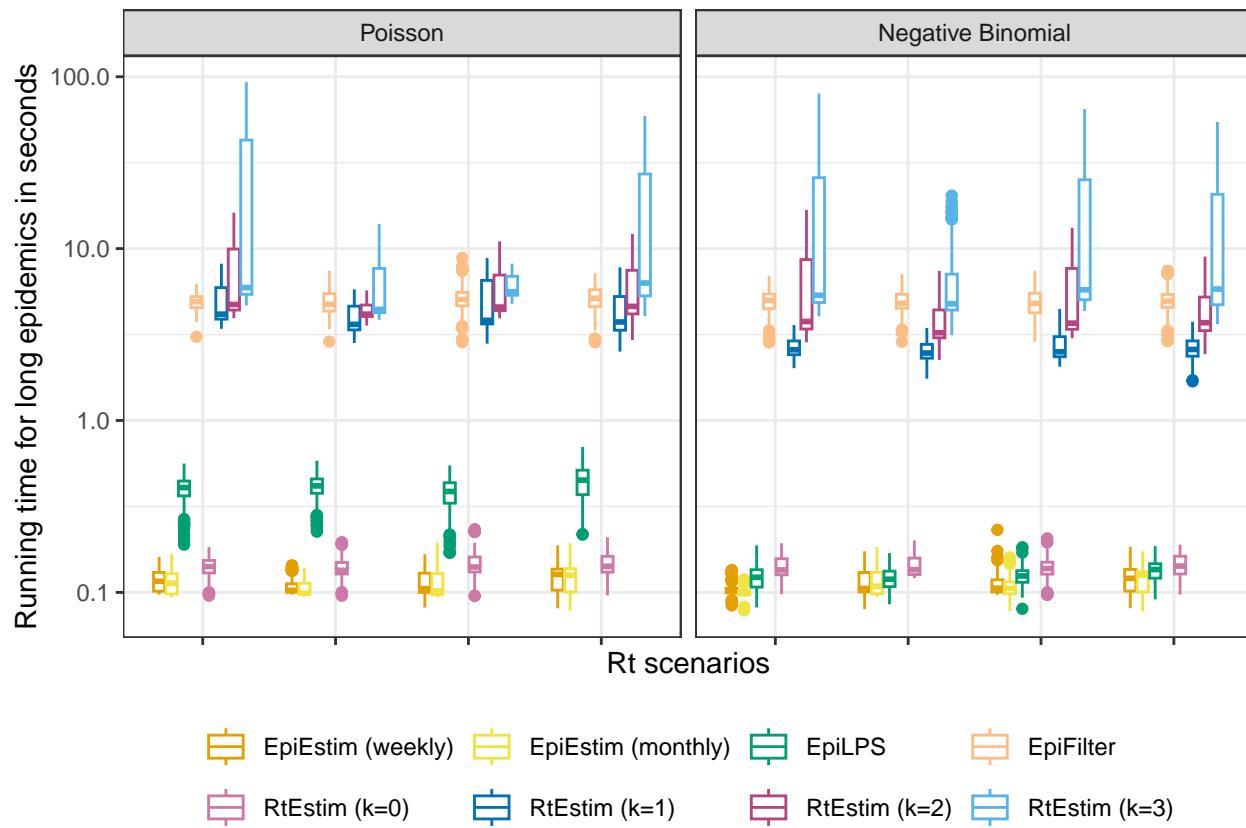


Figure A.6.1: Time comparisons of methods (excluding one outlier of ‘RtEstim (k=1)’ in Scenario 2 with negative Binomial incidence). Y-axis is on a logarithmic scale.

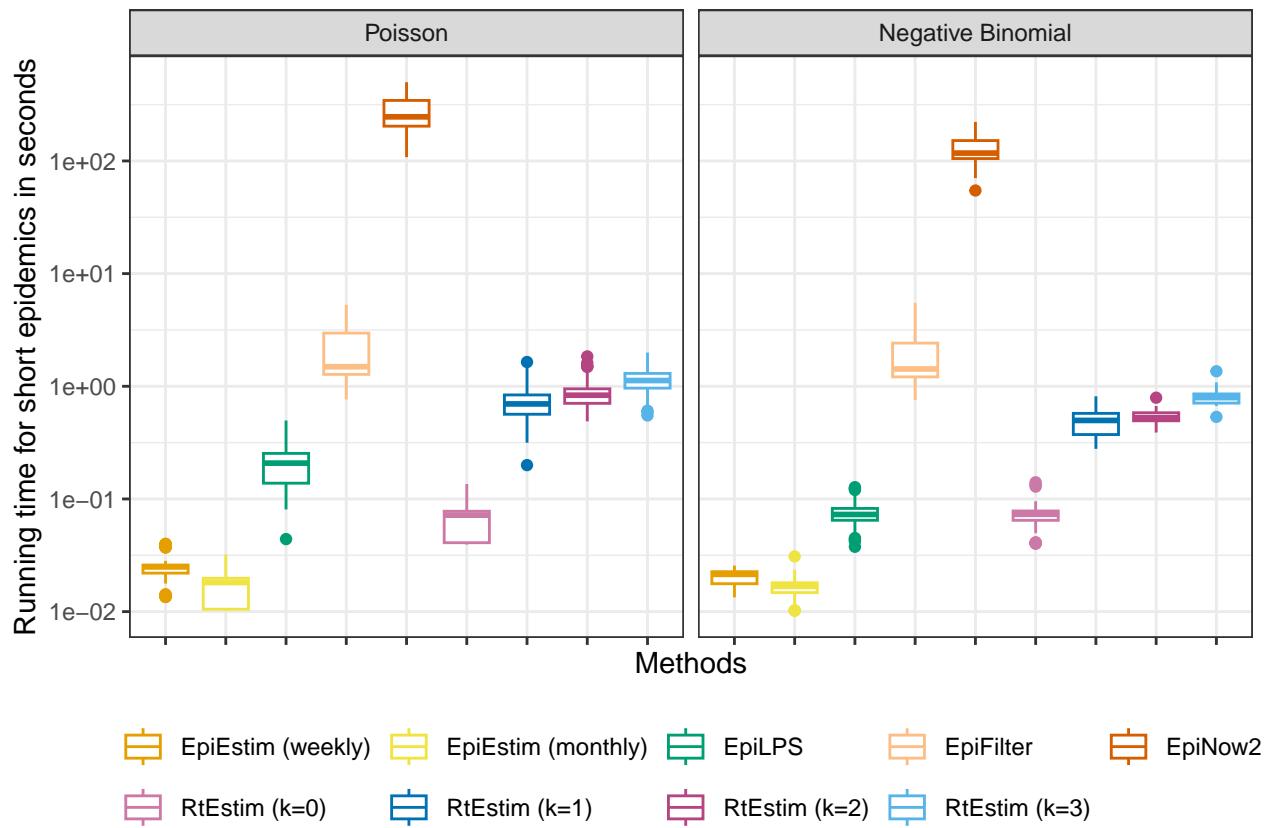


Figure A.6.2: Time comparisons of methods (excluding one outlier of ‘RtEstim (k=1)’ in Scenario 2 with negative Binomial incidence). Y-axis is on a logarithmic scale.