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

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

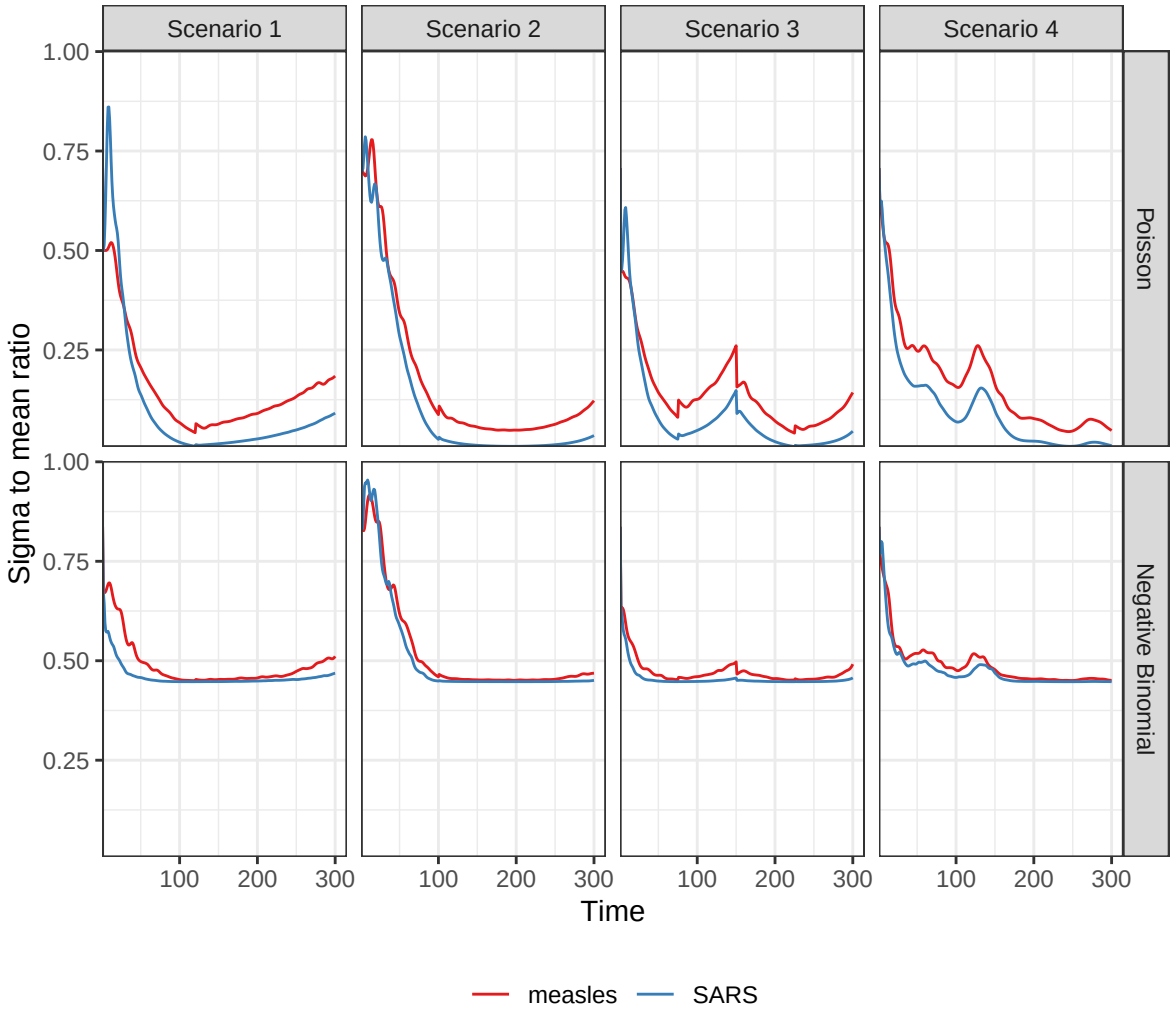
1 Supplmentary details on experimental settings

We compare the accuracy of effective reproduction number estimation using Kullback divergence averaged per coordinate across EpiEstim with weekly and monthly sliding windows, EpiLPS, EpiFilter, EpiNow2, and our RtEstim with k=0,1,2,3. We consider two length of epidemics, $n = 50, 300$. Since EpiNow2 runs too long (specifically, for one epidemic with n=300, it takes ???TBA), we exclude it from the comparison for long epidemics.

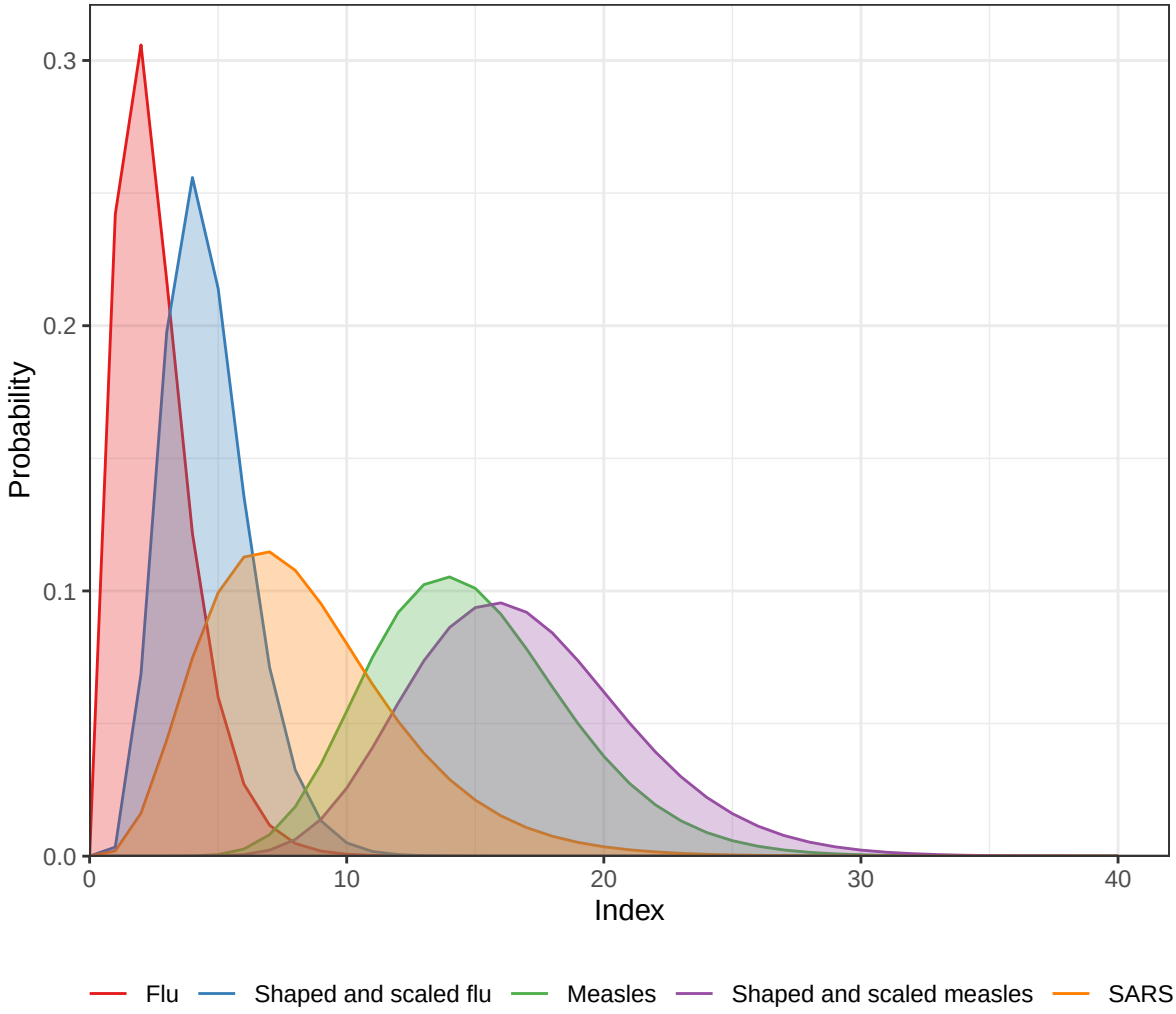
We consider serial interval (SI) distributions of measles and SARS to generate long synthetic epidemics, and flu for short epidemics. Incident cases in synthetic measles epidemics are relatively low (within 1000 at the peak overall), and incidence in SARS is relatively large (between 15000 and 20000 at the peak overall). We specifically consider a reasonably large overdispersion level of negative Binomial incidence is of size 5. Figure @ref(fig:NB-overdispersion) displays the signal to noise ratio (SNR) across different settings, and compared to the counterpart of Poisson incidence (roughly fluctuating in $[1, 25]$ for measles and $[1, 140]$ for SARS), the negative Binomial incidence appears to have an apparently smaller SNR (roughly in $[1, 2.5]$ for both epidemics).

In model fitting, we consider to 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 of SI distributions for both short flu and long measles epidemics, denoted as `flu_ss` and `measles_ss` respectively. It results in seven pairs of SI distributions for incidence generating and 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 @ref(fig:si-dist) displays the SI distributions of `measles`, `measles_ss`, `SARS`, `flu`, and `flu_ss`.

For long epidemics with 300 timepoints, we consider 4 SI pairs applied on 2 incidence distributions using 4 $\mathcal{R}_t$ scenarios solved by 8 methods including EpiEstim with weekly and monthly sliding windows, EpiLPS, EpiFilter, and RtEstim with k=0,1,2,3, which results in 256 experimental settings in total. Each experimental setting is replicated for 50 times, which gives 12800 experiments. For short epidemics with 50 timepoints, we consider 3 SI pairs with 2 incidence distributions and 1 $\mathcal{R}_t$ scenario, solved by 9 methods including the methods for long epidemics and EpiNow2, which results in 54 experimental settings. With 50 replicates, there are 2700 experiments.



Dispersion/Overdispersion level of incidence


0.3
0.2
0.1
0.0
Probability
0
10
20
30
40
Index
Flu
Shaped and scaled flu
Measles
Shaped and scaled measles
SARS

SI curves

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 Rt with mean 1/2000, and use the smoothed Rt as estimates. We run 10-fold cross validation (CV) to choose the best tuning parameter from the candidate set of size 50, i.e., $\boldsymbol{\lambda} = \{\lambda_1, \cdots, \lambda_{50}\}$. 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 have visualized the main results of accuracy (using KL excluding the first week) for long epidemics under all settings across all methods in the section of results for synthetic data. We will visualize other main experimental results in the following.

2 Supplementary experimental results on long epidemics for Section 3.1

Fig 10 displays the KL divergence values for negative Binomial incidence. Comparing across $\mathcal{R}_t$ estimates by EpiLPS, RtEstim and EpiEstim with ***monthly*** sliding windows, KL divergence computation excludes the first month of $\mathcal{R}_t$ estimates for all approaches, since EpiEstim estimates with the monthly sliding windows are not available until the second month. 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.

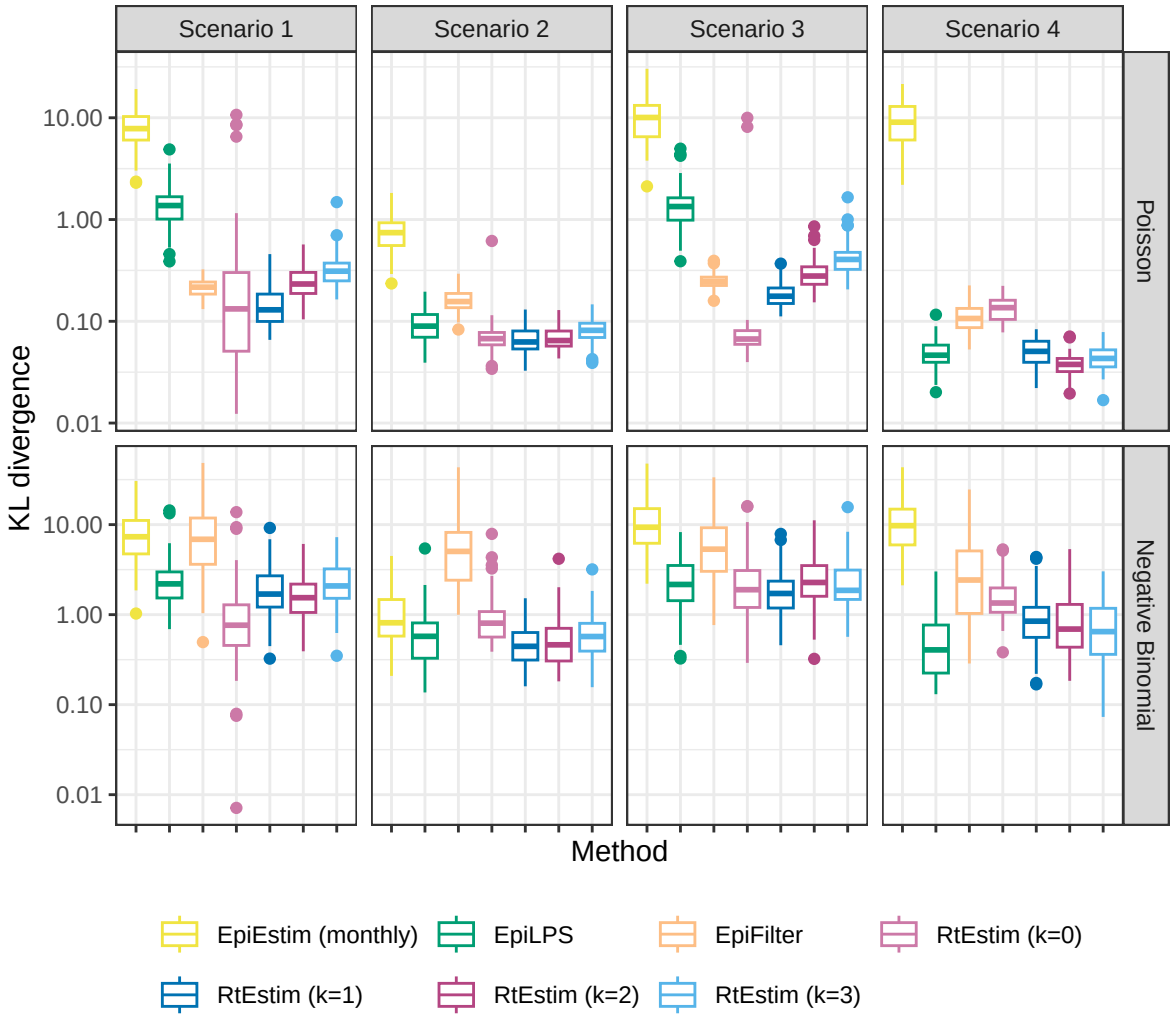
3 Experimental results on short epidemics

4 Misspecification of serial interval function parameters

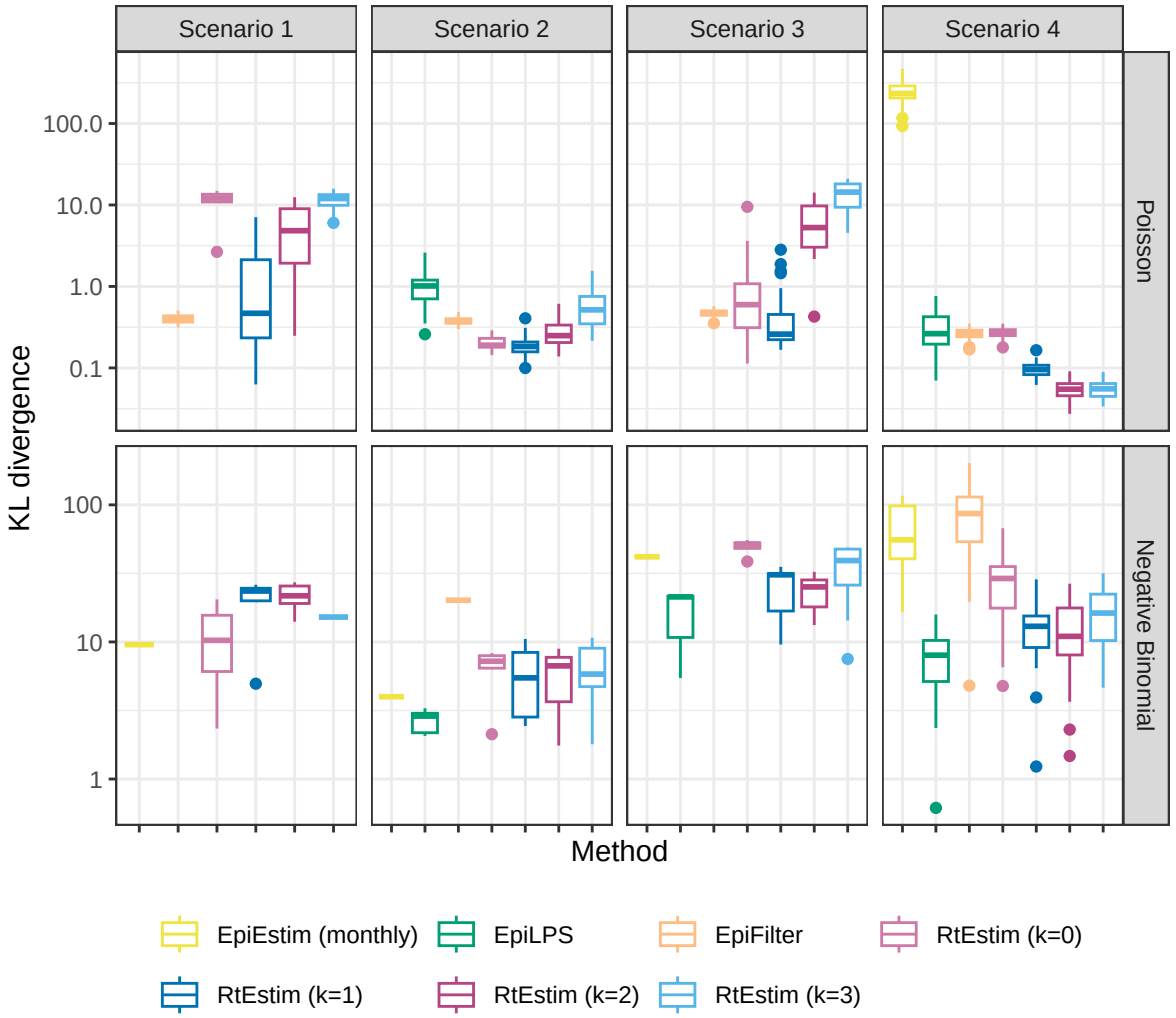
4.1 Mild misspecification

For long epidemics, ...

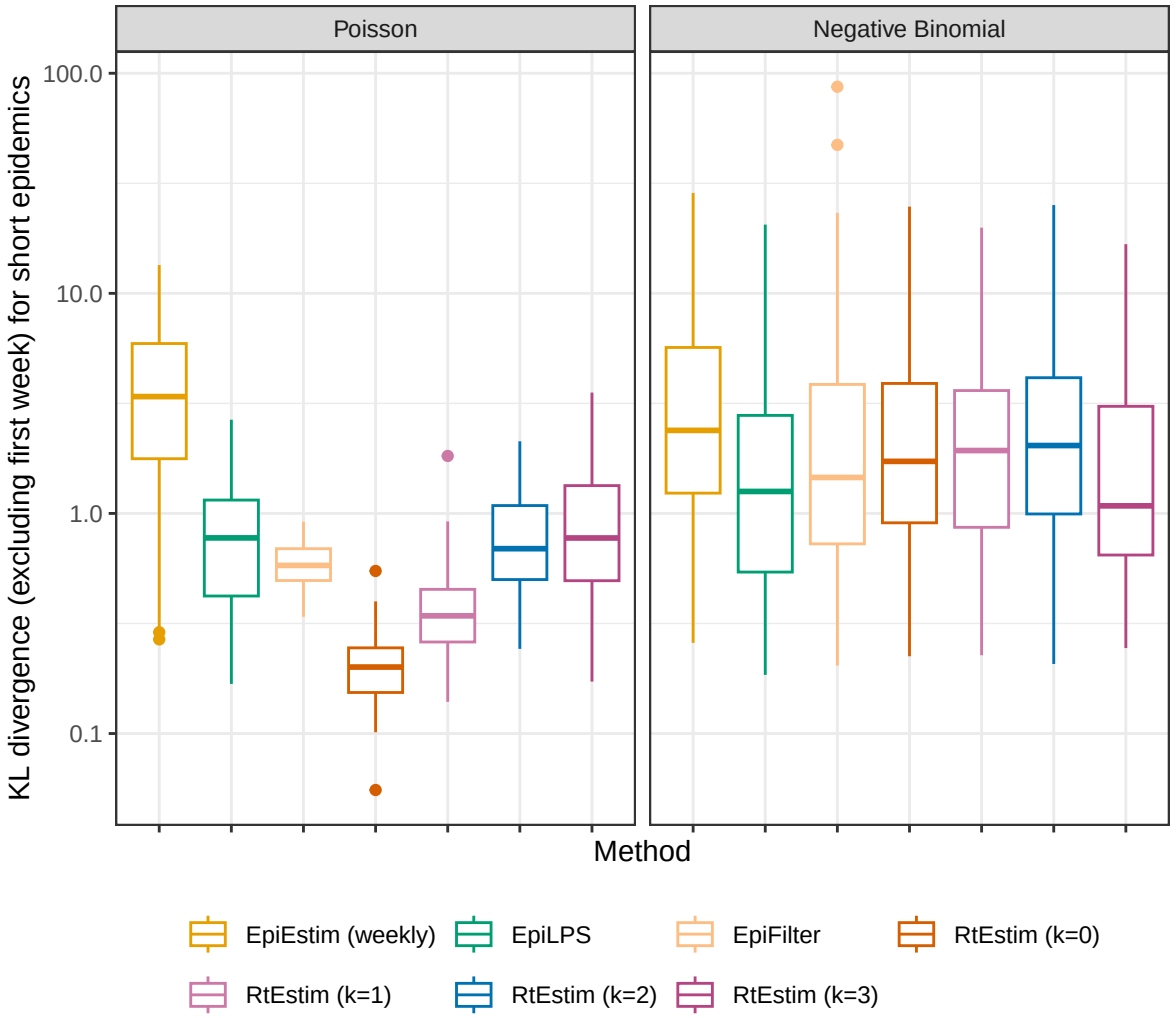
For short epidemics,



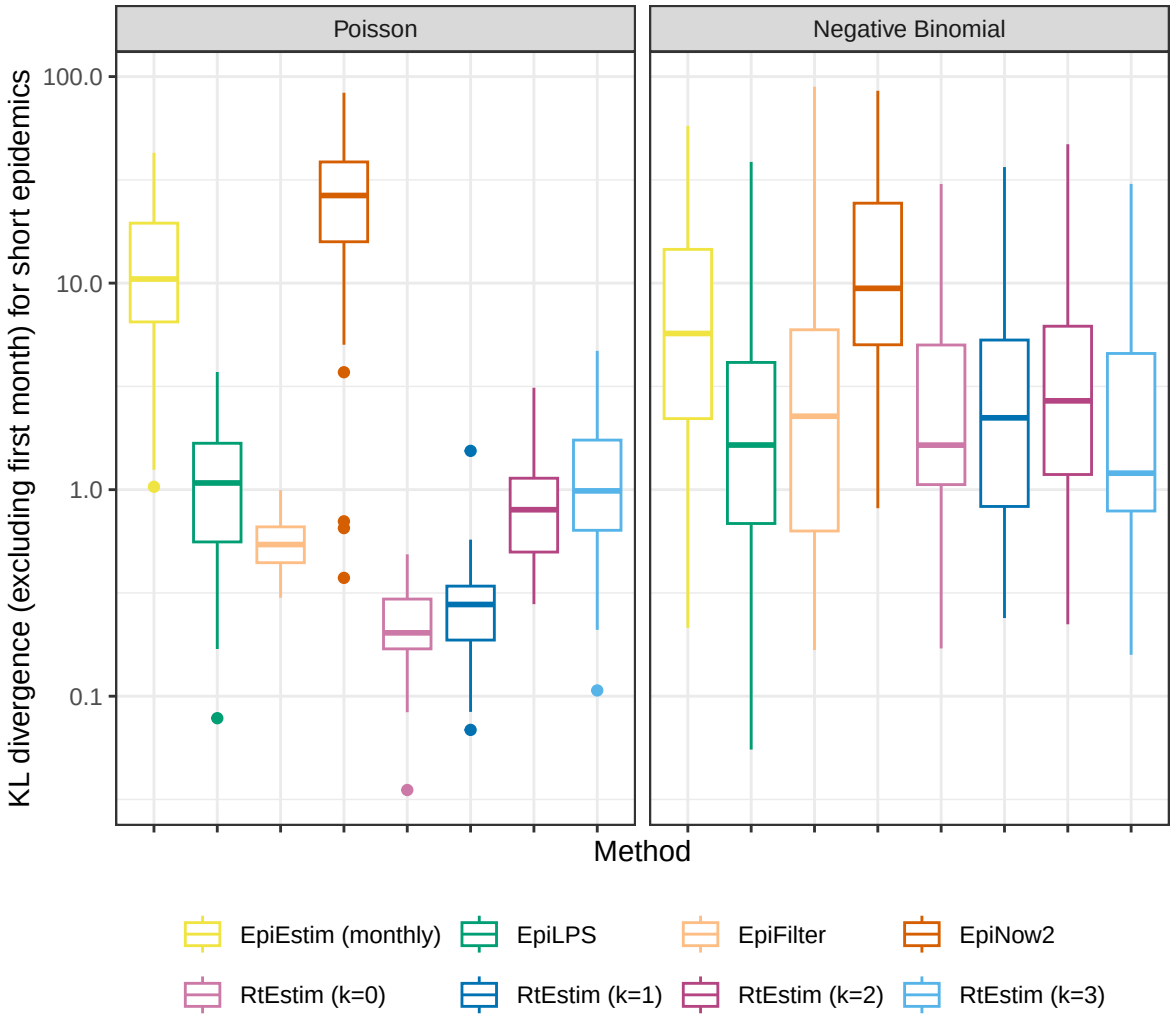
KL divergence excluding the first months for measles epidemics. Y-axis is on a logarithmic scale.



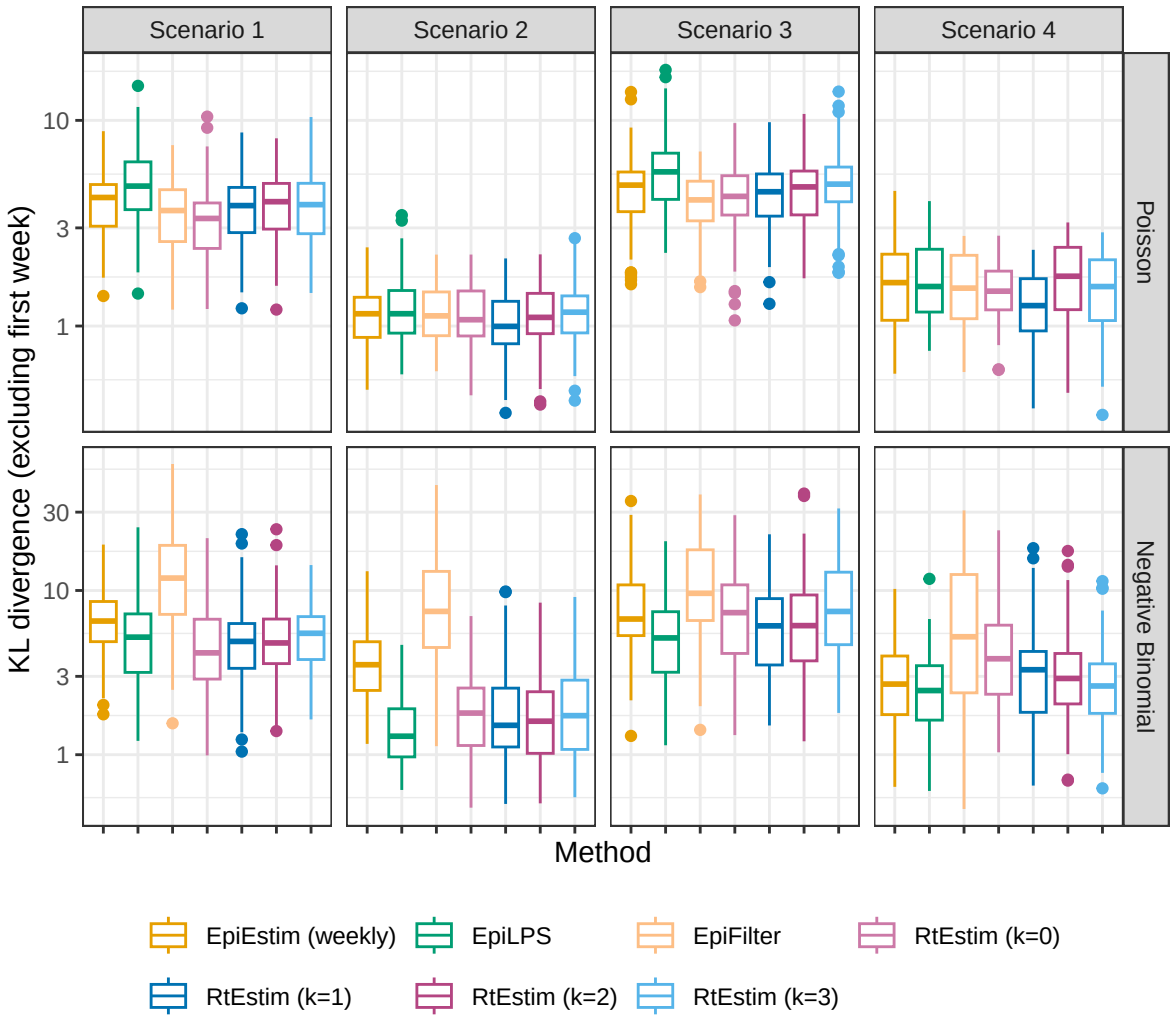
KL divergence excluding the first months for SARS epidemics. Y-axis is on a logarithmic scale.



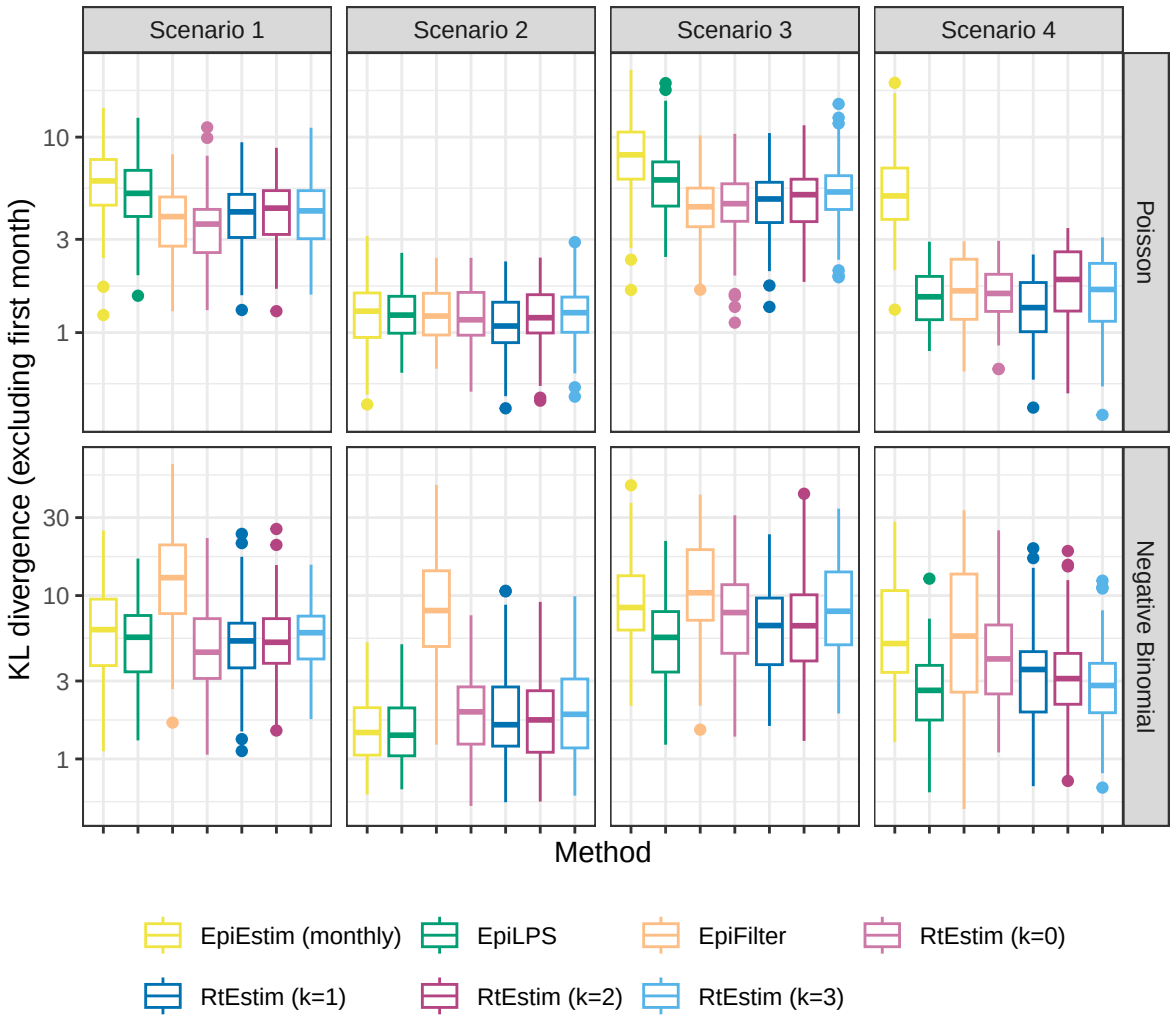
KL divergence excluding the first weeks for flu epidemics. Y-axis is on a logarithmic scale.



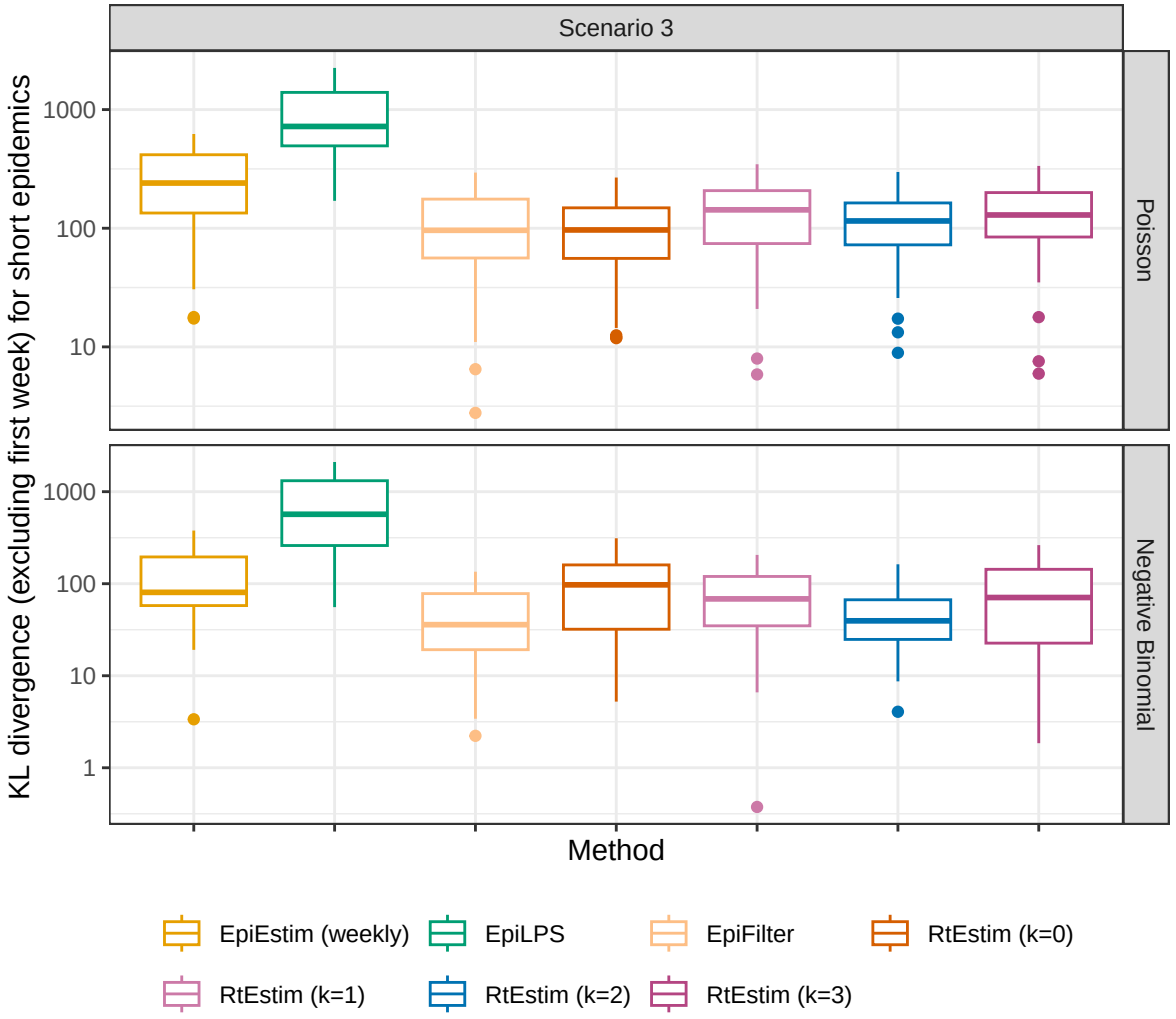
KL divergence excluding the first months for flu epidemics. Y-axis is on a logarithmic scale.



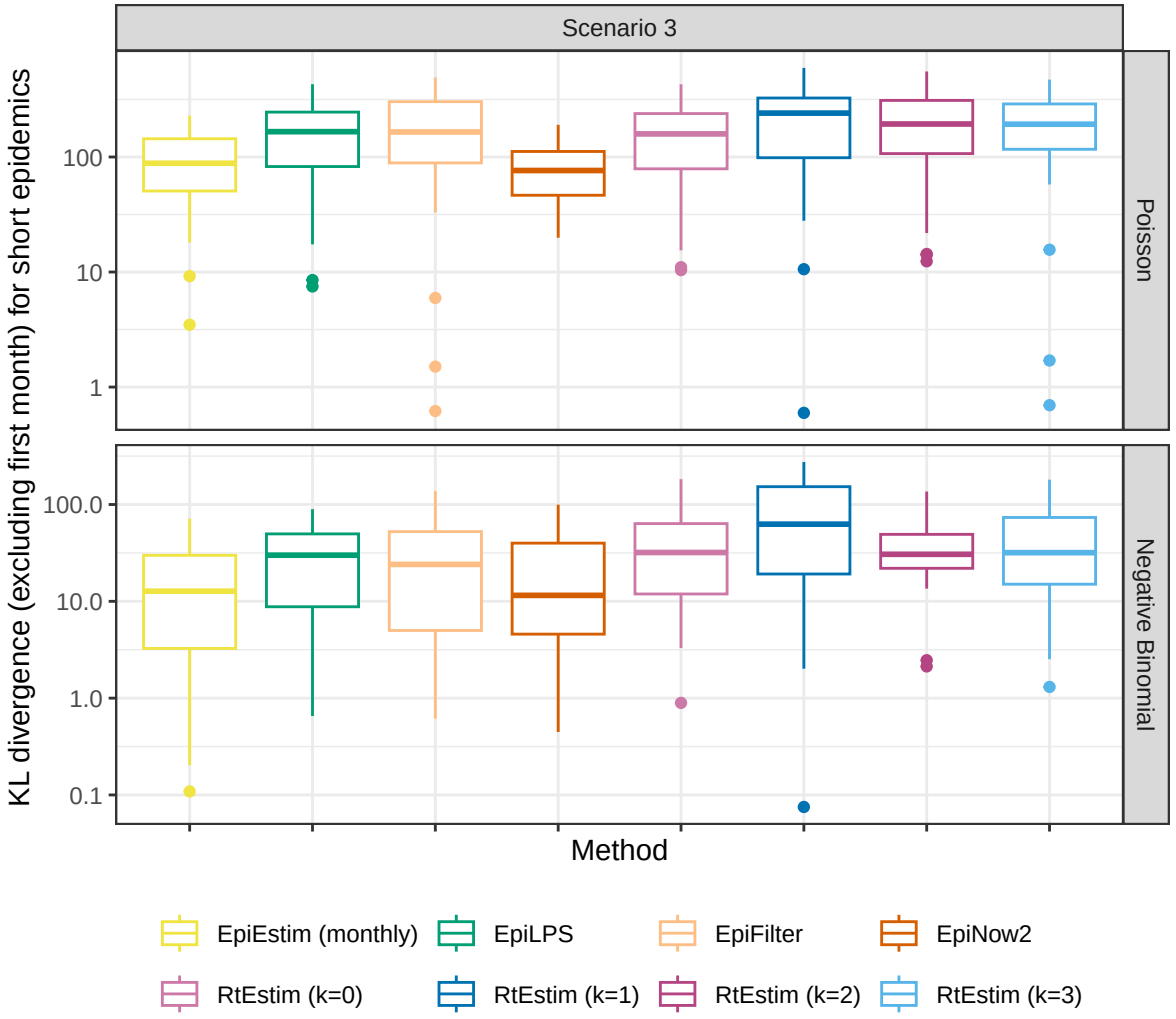
KL divergence excluding the first weeks for measle epidemics with major SI misspecification. Y-axis is on a logarithmic scale.



KL divergence excluding the first months for measle epidemics with major SI misspecification. Y-axis is on a logarithmic scale.



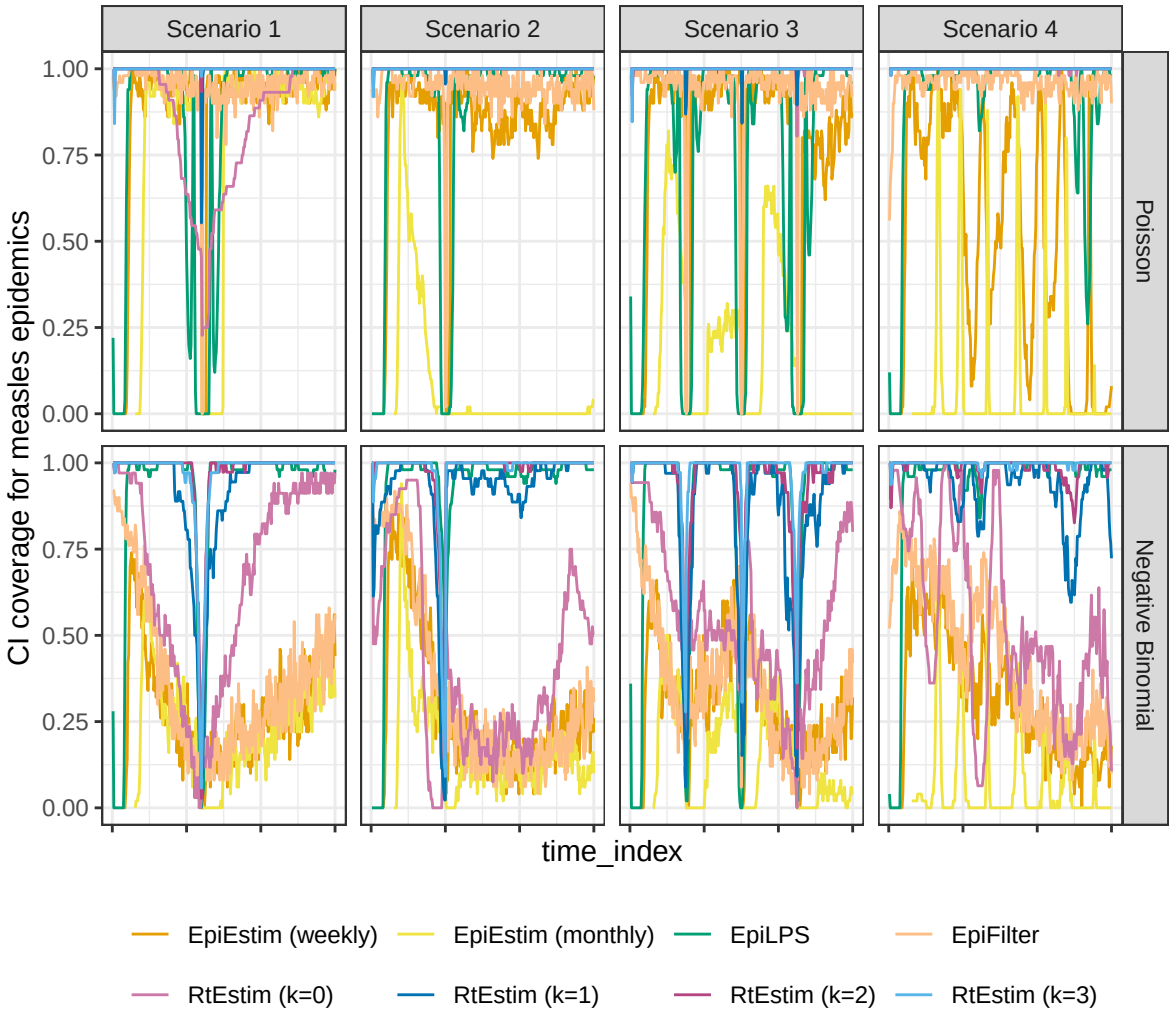
KL divergence excluding the first weeks for flu epidemics with major SI misspecification. Y-axis is on a logarithmic scale.

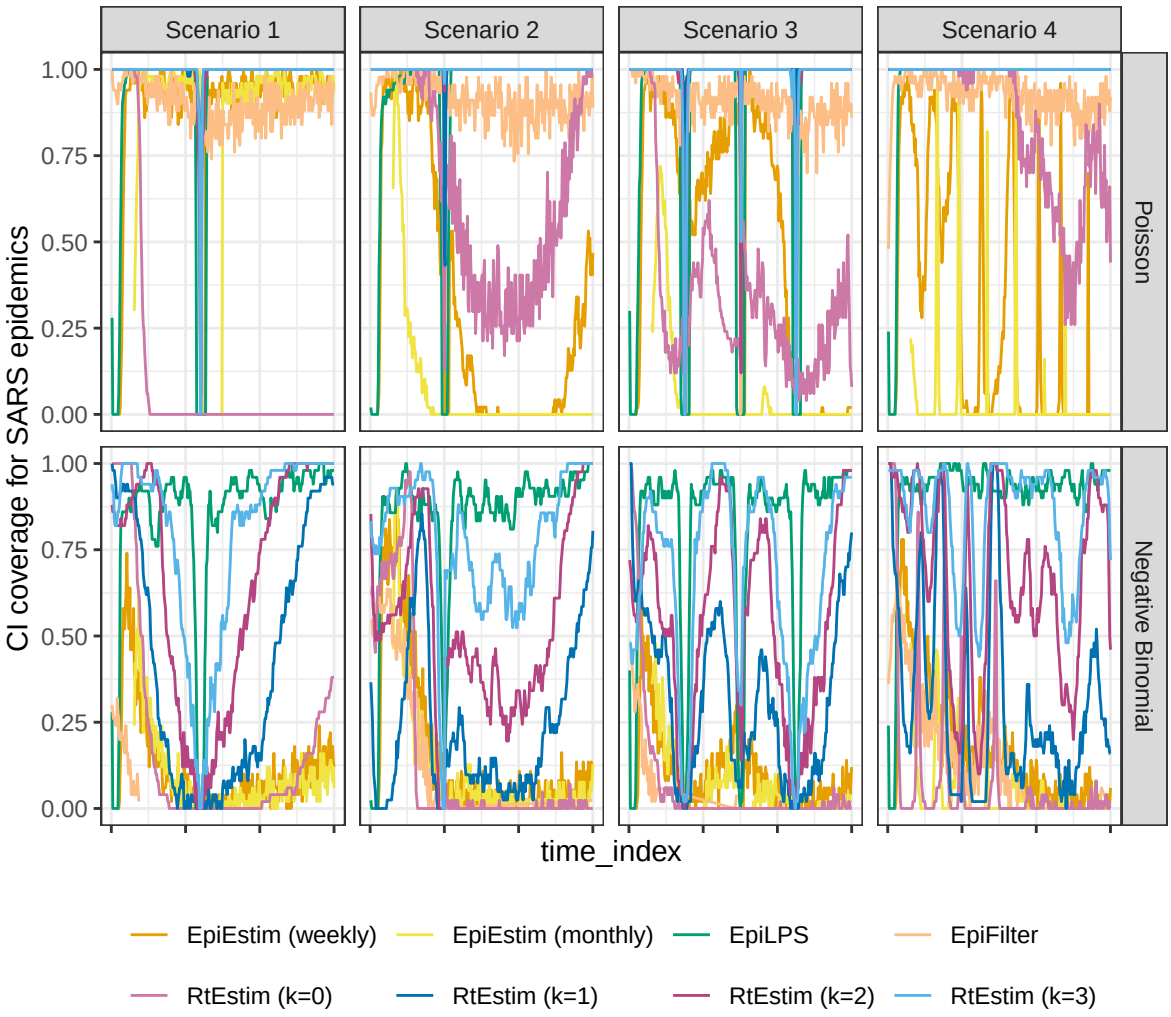


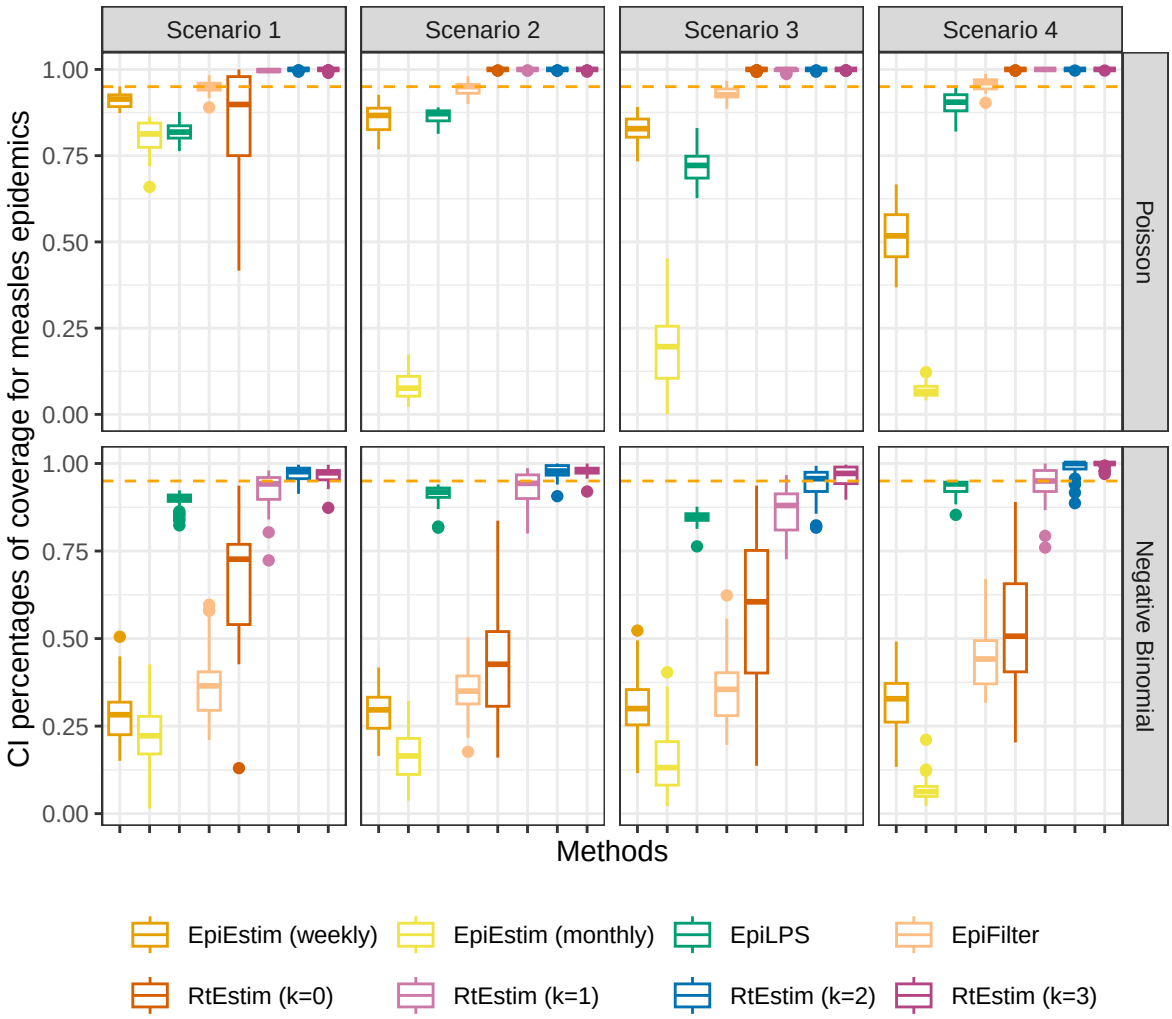
KL divergence excluding the first months for flu epidemics with major SI misspecification. Y-axis is on a logarithmic scale.

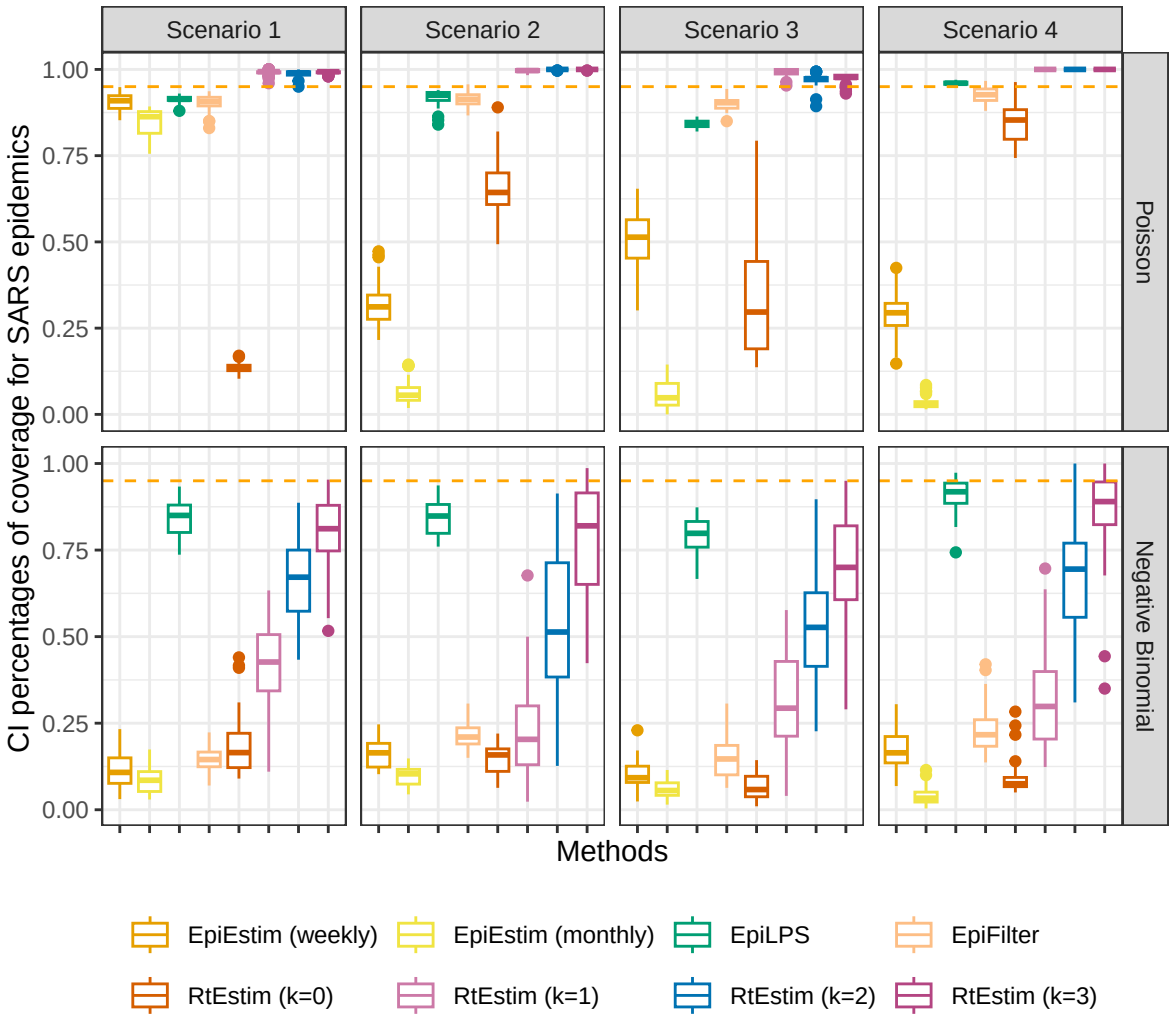
4.2 Major misspecification

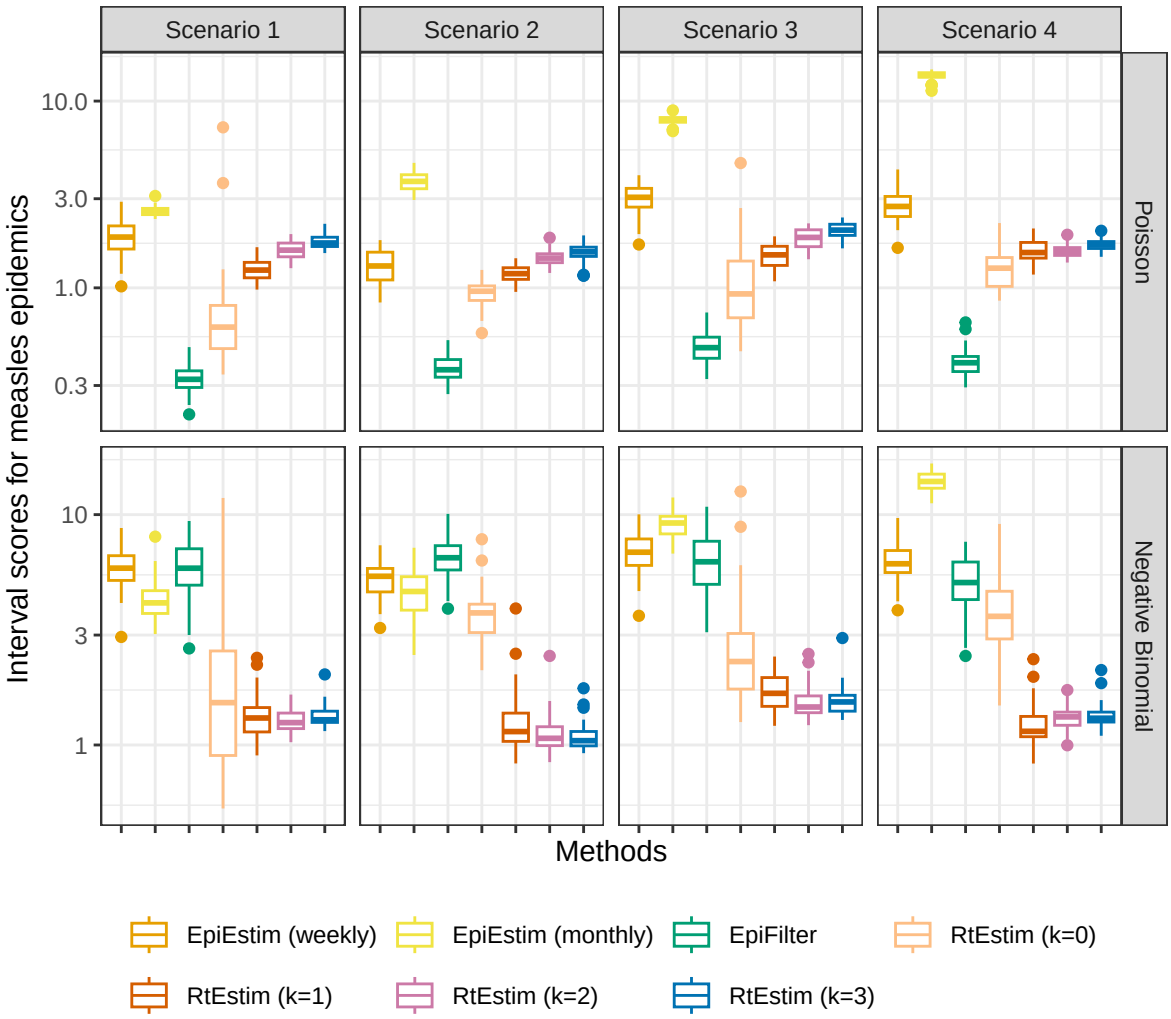
5 Confidence interval coverage

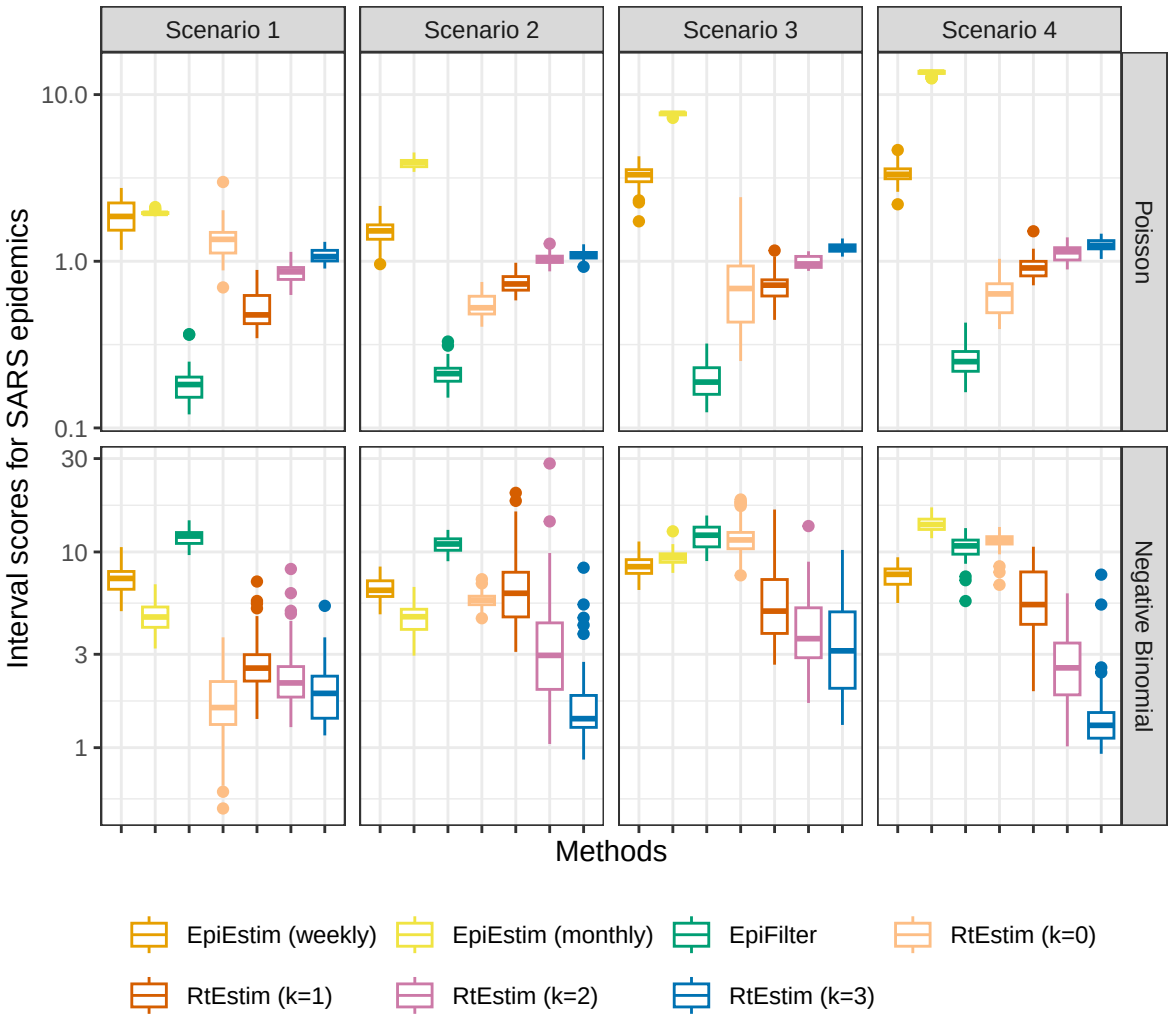



Scenario 1
Scenario 2
Scenario 3
Scenario 4
Poisson
Negative Binomial
CI coverage for SARS epidemics
1.00
0.75
0.50
0.25
0.00
time_index
EpiEstim (weekly)
EpiEstim (monthly)
EpiLPS
EpiFilter
RtEstim (k=0)
RtEstim (k=1)
RtEstim (k=2)
RtEstim (k=3)


Scenario 1
Scenario 2
Scenario 3
Scenario 4
Poisson
Negative Binomial
CI percentages of coverage for measles epidemics
1.00
0.75
0.50
0.25
0.00
Methods
EpiEstim (weekly)
EpiEstim (monthly)
EpiLPS
EpiFilter
RtEstim (k=0)
RtEstim (k=1)
RtEstim (k=2)
RtEstim (k=3)


Scenario 1
Scenario 2
Scenario 3
Scenario 4
Poisson
Negative Binomial
CI percentages of coverage for SARS epidemics
1.00
0.75
0.50
0.25
0.00
Methods
EpiEstim (weekly)
EpiEstim (monthly)
EpiLPS
EpiFilter
RtEstim (k=0)
RtEstim (k=1)
RtEstim (k=2)
RtEstim (k=3)

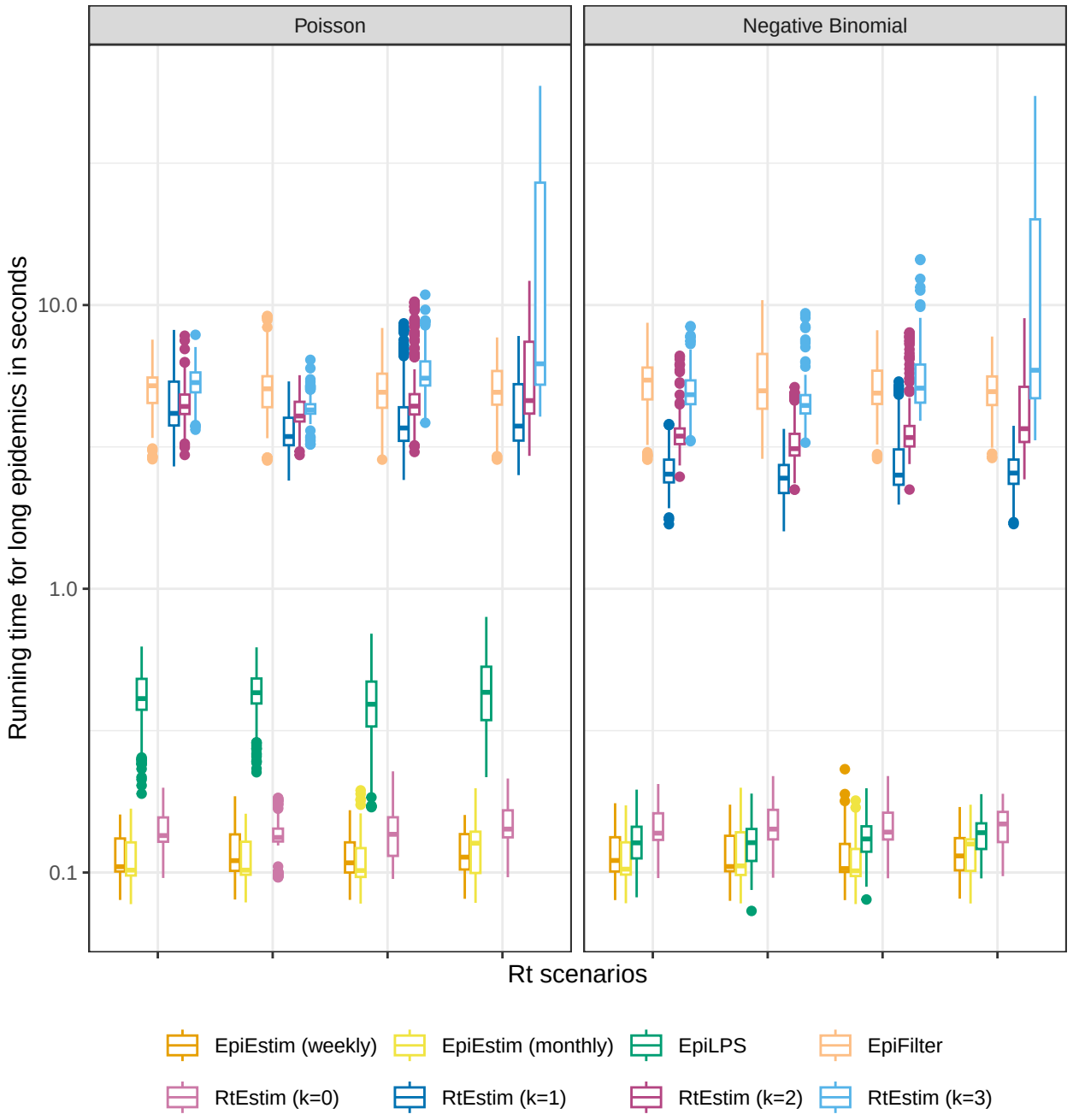

Scenario 1
Scenario 2
Scenario 3
Scenario 4
Poisson
Negative Binomial
Interval scores for measles epidemics
10.0
3.0
1.0
0.3
10
3
1
Methods
EpiEstim (weekly)
EpiEstim (monthly)
EpiFilter
RtEstim (k=0)
RtEstim (k=1)
RtEstim (k=2)
RtEstim (k=3)

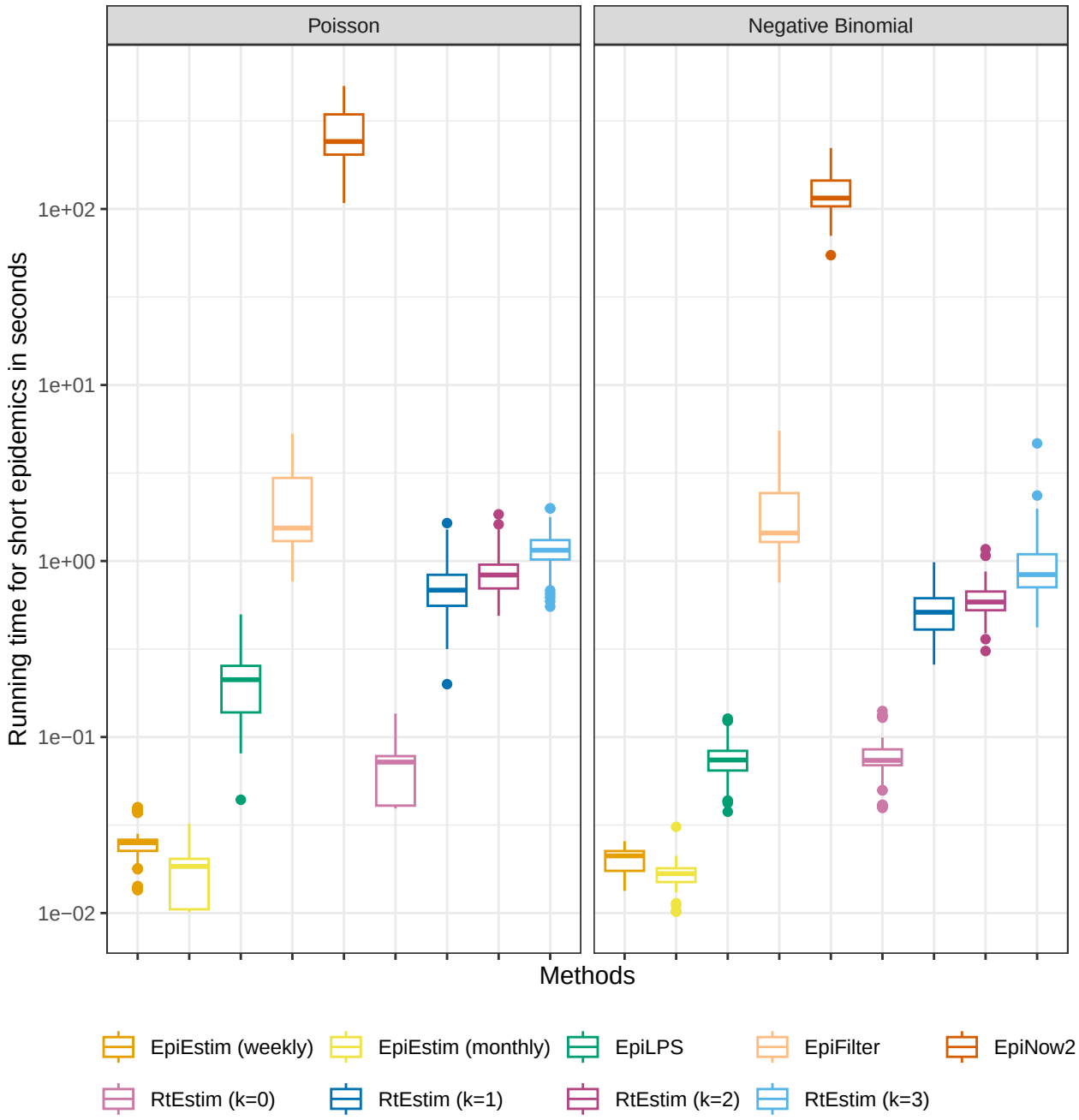


6 Time comparisons of methods for Section 3.2

Fig 12 shows 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.


Poisson
Negative Binomial
Running time for long epidemics in seconds
10.0
1.0
0.1
Rt scenarios
EpiEstim (weekly)
EpiEstim (monthly)
EpiLPS
EpiFilter
RtEstim (k=0)
RtEstim (k=1)
RtEstim (k=2)
RtEstim (k=3)


Poisson
Negative Binomial
Running time for short epidemics in seconds
1e+02
1e+01
1e+00
1e−01
1e−02
Methods
EpiEstim (weekly)
EpiEstim (monthly)
EpiLPS
EpiFilter
EpiNow2
RtEstim (k=0)
RtEstim (k=1)
RtEstim (k=2)
RtEstim (k=3)