Supplementary Materials S2: Additional Simulation Studies

Regression Models for COVID-19 Epidemic Dynamics with Incomplete Data

Corbin Quick, Rounak Dey, and Xihong Lin

June 1, 2021

In this document, we provide more comprehensive figures for 3 sets of simulations: (A) the correctly specified model described in the main text, (B) a second set of correctly specified simulations, and (C) simulations with misspecified nuisance parameters. In (A), we considered 4 settings for the effective reproductive number R_{it} and a single setting for the ascertainment probability π_{it}. In (B), we recycle the 4 specifications for R_{it} from the previous simulations and change the specification of the π_{it} model. In (C), we hold the R_{it} and π_{it} models constant and consider 8 specification of the nuisance parameters, each of which are misspecified.

For each set of simulations, we show results for (1) the effective reproductive number R_{it}, (2) the ascertainment probability π_{it}, (3) the daily incidence (the fraction of the population infected on each day), and (4) the cumulative prevalence (the fraction of the population that has ever been infection). For each category of results, we show the true (data-generating) values and estimates from MERMAID and other models, the bias of MERMAID estimates, and the calibration of MERMAID standard error estimates.

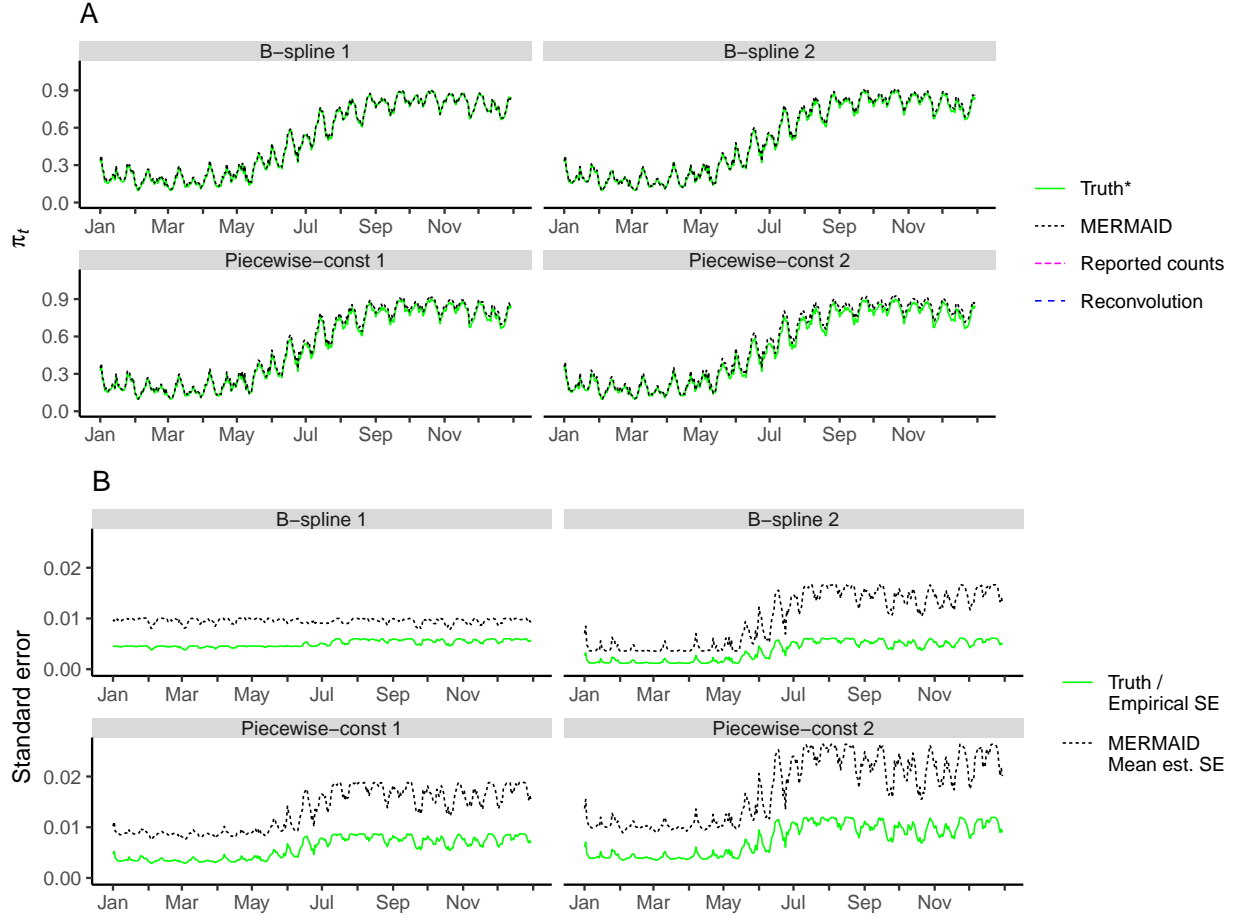# A Correctly Specified Simulations 1

A.1 Effective reproductive numbers R_{it}
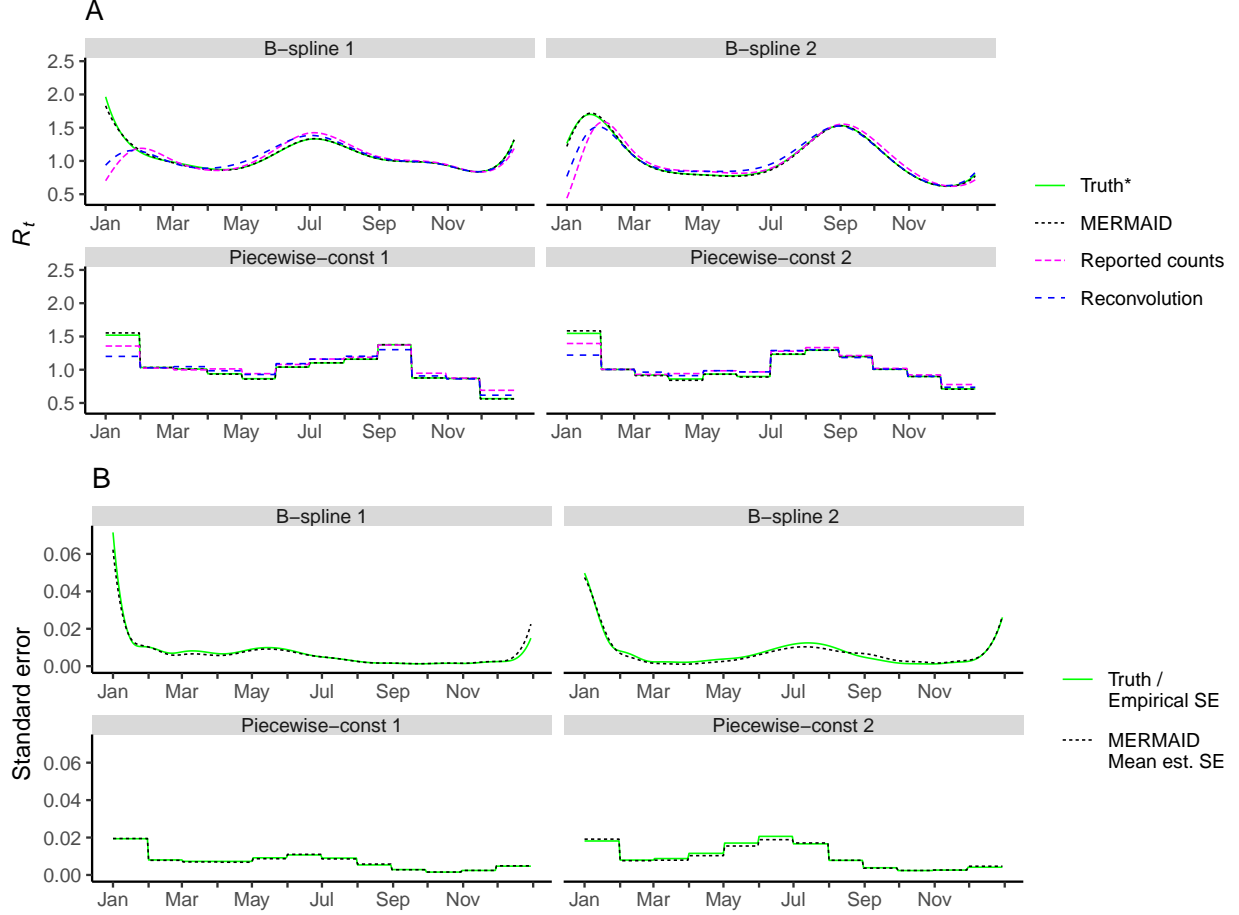
Supplementary Figure 1: Estimates of R_{it} in 4 simulation scenarios under the correctly specified model, where the R_{it} is specified either using B-splines or piece-wise constant functions. **Panel A:** Shown are curves for true R_{it} (solid green), and $\hat{R}_{it}$ estimated using MERMAID (black), GLM with raw reported case counts (magenta), or GLM with reconvoluted case counts (blue). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).

Note: These 4 simulation settings are identical to those shown in the main text, and are reproduced here for completeness.

A.2 Ascertainment probabilities π_{it}

Supplementary Figure 2: Estimates of π_{it} in 4 simulation scenarios under the correctly specified model, where the R_{it} is specified either using B-splines or piece-wise constant functions and π_{it} is constant across the 4 scenarios. **Panel A:** Shown are curves for true π_{it} (solid green), and $\hat{\pi}_{it}$ estimated using MERMAID (black). Here, only MERMAID is shown, as the other methods ignore ascertainment. **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates). Here, estimated standard errors are conservative due to the approximate correction for seroprevalence likelihood weights described in the main text.

Note: These 4 simulation settings are identical to those shown in the main text, and are reproduced here for completeness.
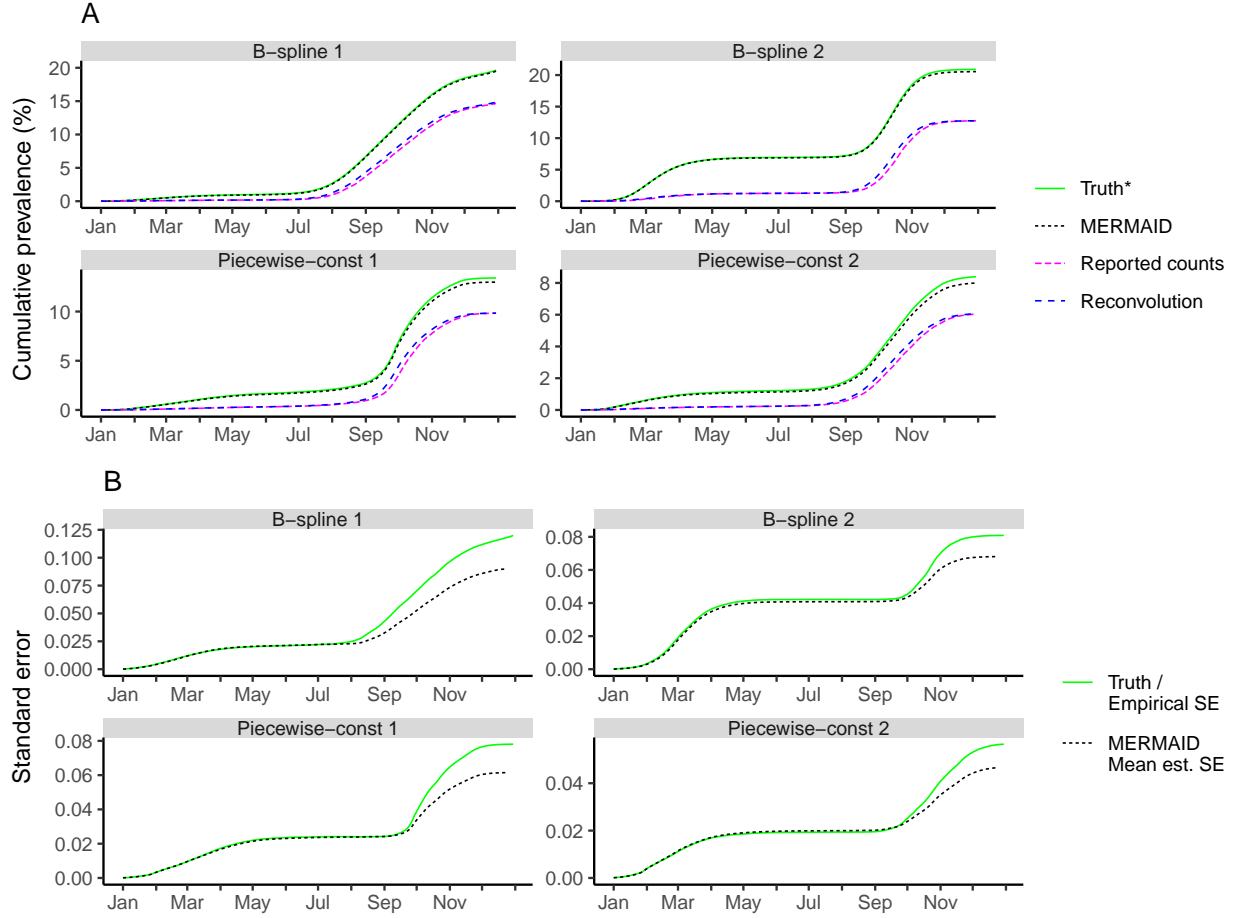
A.3 Daily incidence
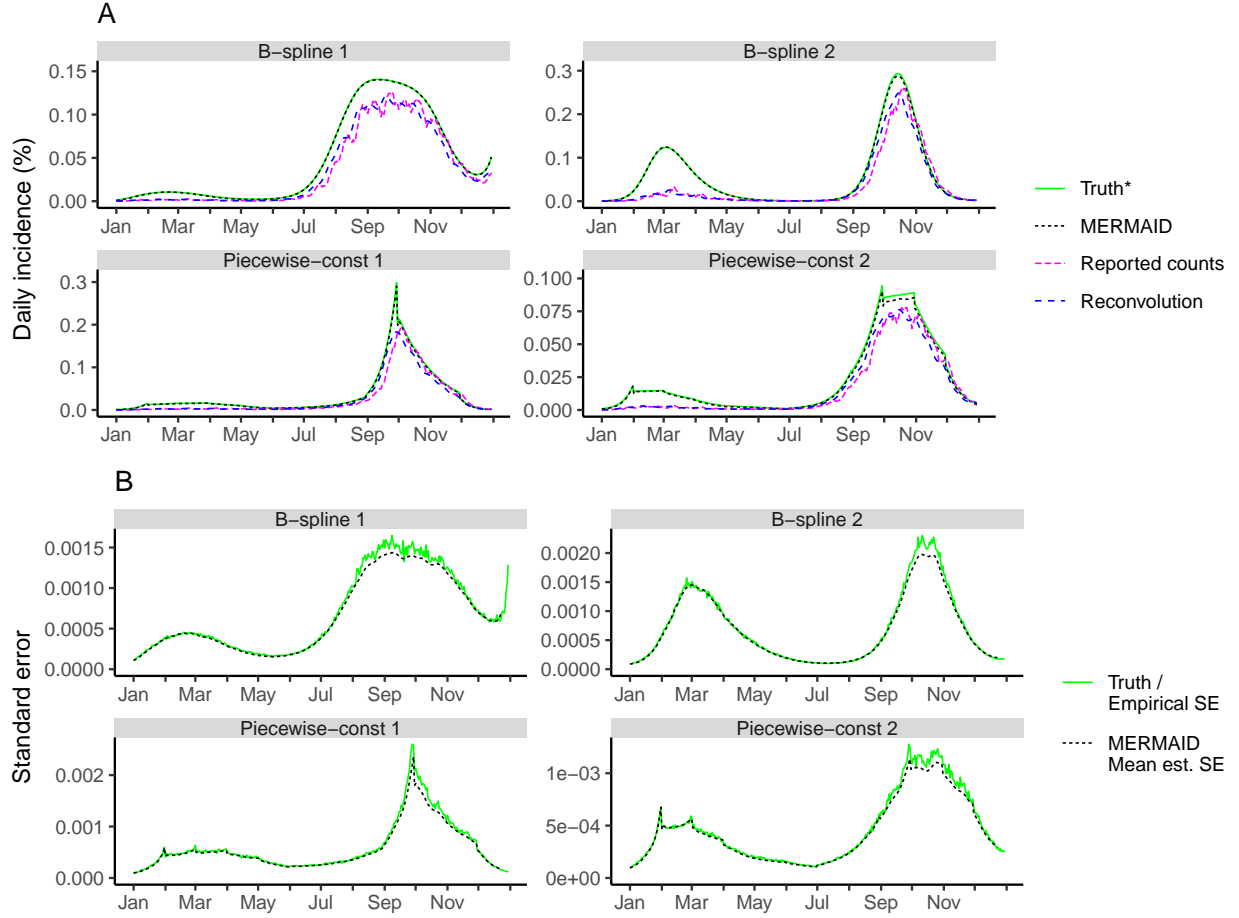
Supplementary Figure 3: Estimates of daily incidence in 4 simulation scenarios under the correctly specified model, where the R_{it} is specified either using B-splines or piece-wise constant functions and π_{it} is constant across the 4 scenarios. **Panel A:** Shown are curves for true daily incidence (solid green), and estimates from MERMAID (black), GLM with raw reported case counts (magenta), or GLM with reconvoluted case counts (blue). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).

Note: These 4 simulation settings are identical to those shown in the main text, and are reproduced here for completeness.
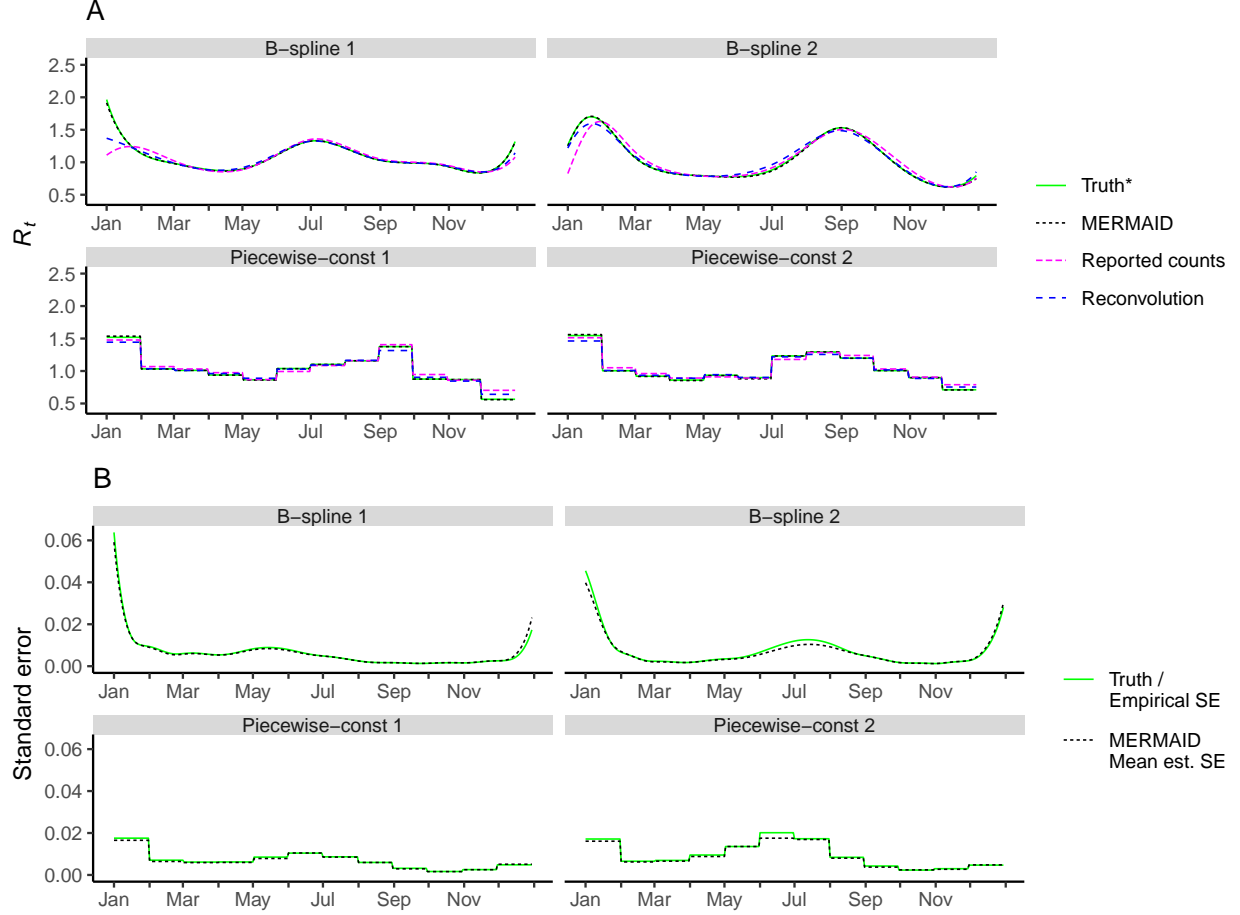
A.4 Cumulative prevalence

Supplementary Figure 4: Estimates of cumulative prevalence in 4 simulation scenarios under the correctly specified model, where the R_{it} is specified either using B-splines or piece-wise constant functions and π_{it} is constant across the 4 scenarios. **Panel A:** Shown are curves for true cumulative prevalence (solid green), and estimates from MERMAID (black), GLM with raw reported case counts (magenta), or GLM with reconvoluted case counts (blue). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).

Note: These 4 simulation settings are identical to those shown in the main text, and are reproduced here for completeness.
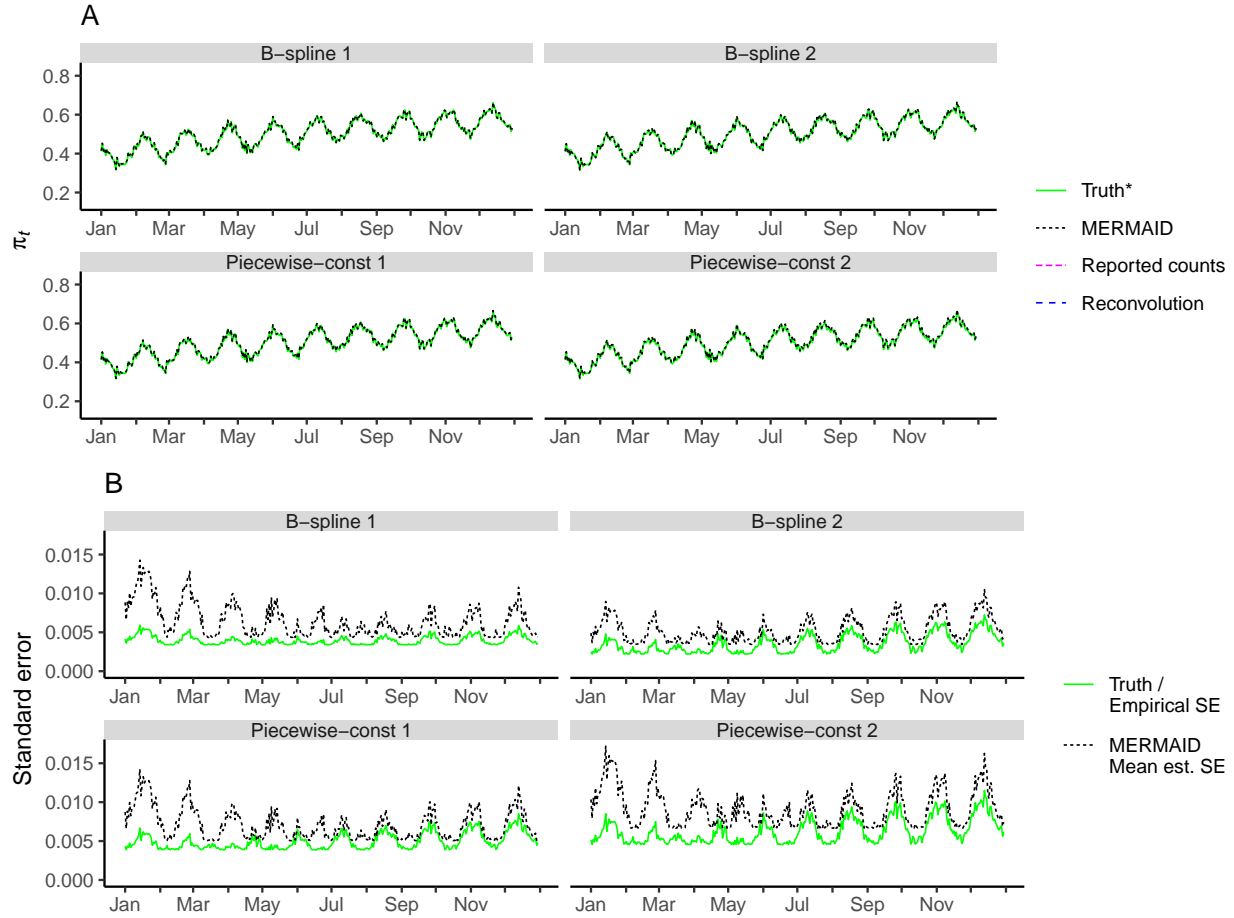
B Correctly Specified Simulations 2

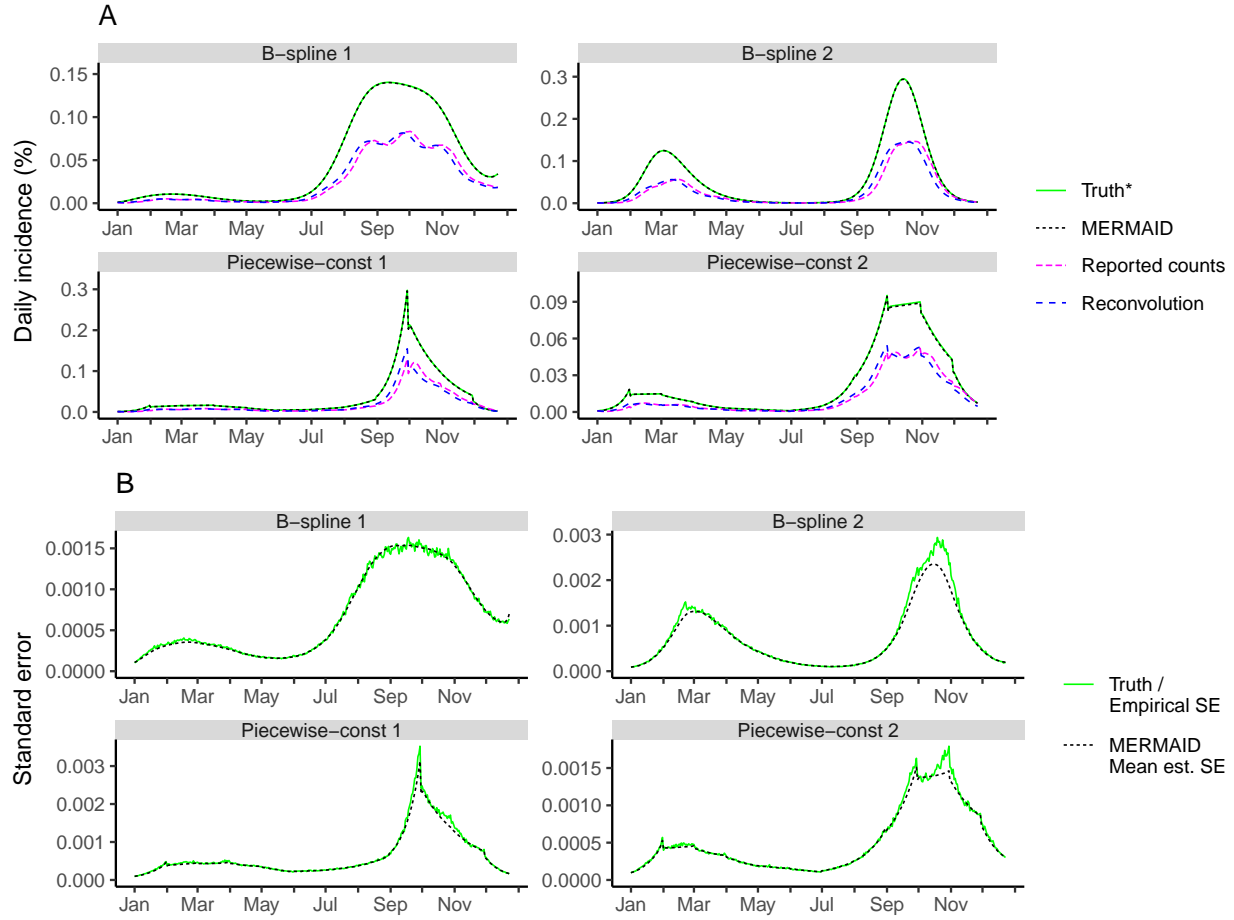## B.1 Effective reproductive numbers R_{it}

Supplementary Figure 5: Estimates of R_{it} in 4 simulation scenarios under the correctly specified model, where the R_{it} is specified either using B-splines or piece-wise constant functions. **Panel A:** Shown are curves for true R_{it} (solid green), and $\hat{R}_{it}$ estimated using MERMAID (black), GLM with raw reported case counts (magenta), or GLM with reconvoluted case counts (blue). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).

B.2 Ascertainment probabilities π_{it}

Supplementary Figure 6: Estimates of π_{it} in 4 simulation scenarios under the correctly specified model, where the R_{it} is specified either using B-splines or piece-wise constant functions and π_{it} is constant across the 4 scenarios. **Panel A:** Shown are curves for true π_{it} (solid green), and $\hat{\pi}_{it}$ estimated using MERMAID (black). Here, only MERMAID is shown, as the other methods ignore ascertainment. **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates). Here, estimated standard errors are conservative due to the approximate correction for seroprevalence likelihood weights described in the main text.
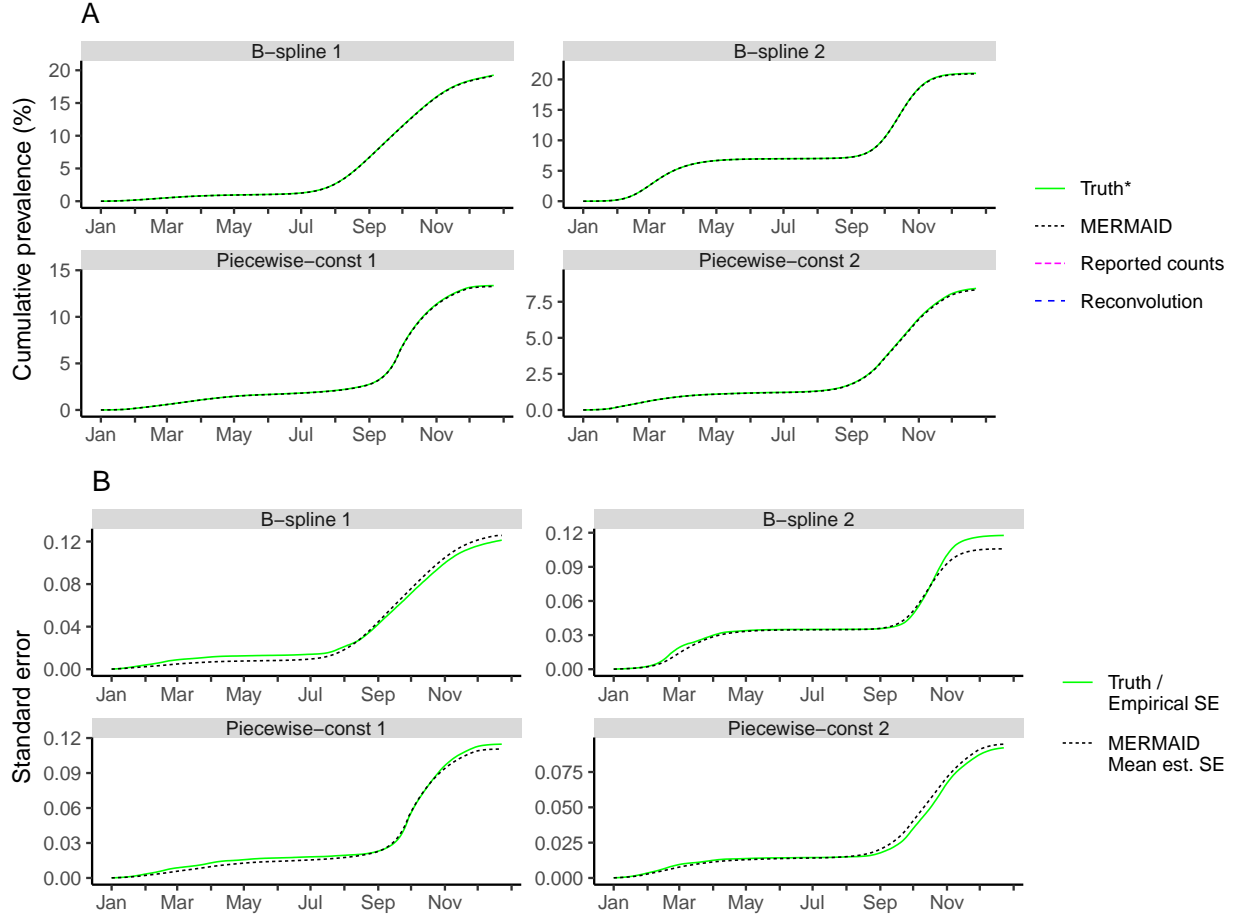
B.3   Daily incidence

Supplementary Figure 7: Estimates of daily incidence in 4 simulation scenarios under the correctly specified model, where the R_{it} is specified either using B-splines or piece-wise constant functions and π_{it} is constant across the 4 scenarios. **Panel A:** Shown are curves for true daily incidence (solid green), and estimates from MERMAID (black), GLM with raw reported case counts (magenta), or GLM with reconvoluted case counts (blue). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).

B.4 Cumulative prevalence

Supplementary Figure 8: Estimates of cumulative prevalence in 4 simulation scenarios under the correctly specified model, where the R_{it} is specified either using B-splines or piece-wise constant functions and π_{it} is constant across the 4 scenarios. **Panel A:** Shown are curves for true cumulative prevalence (solid green), and estimates from MERMAID (black), GLM with raw reported case counts (magenta), or GLM with reconvoluted case counts (blue). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).
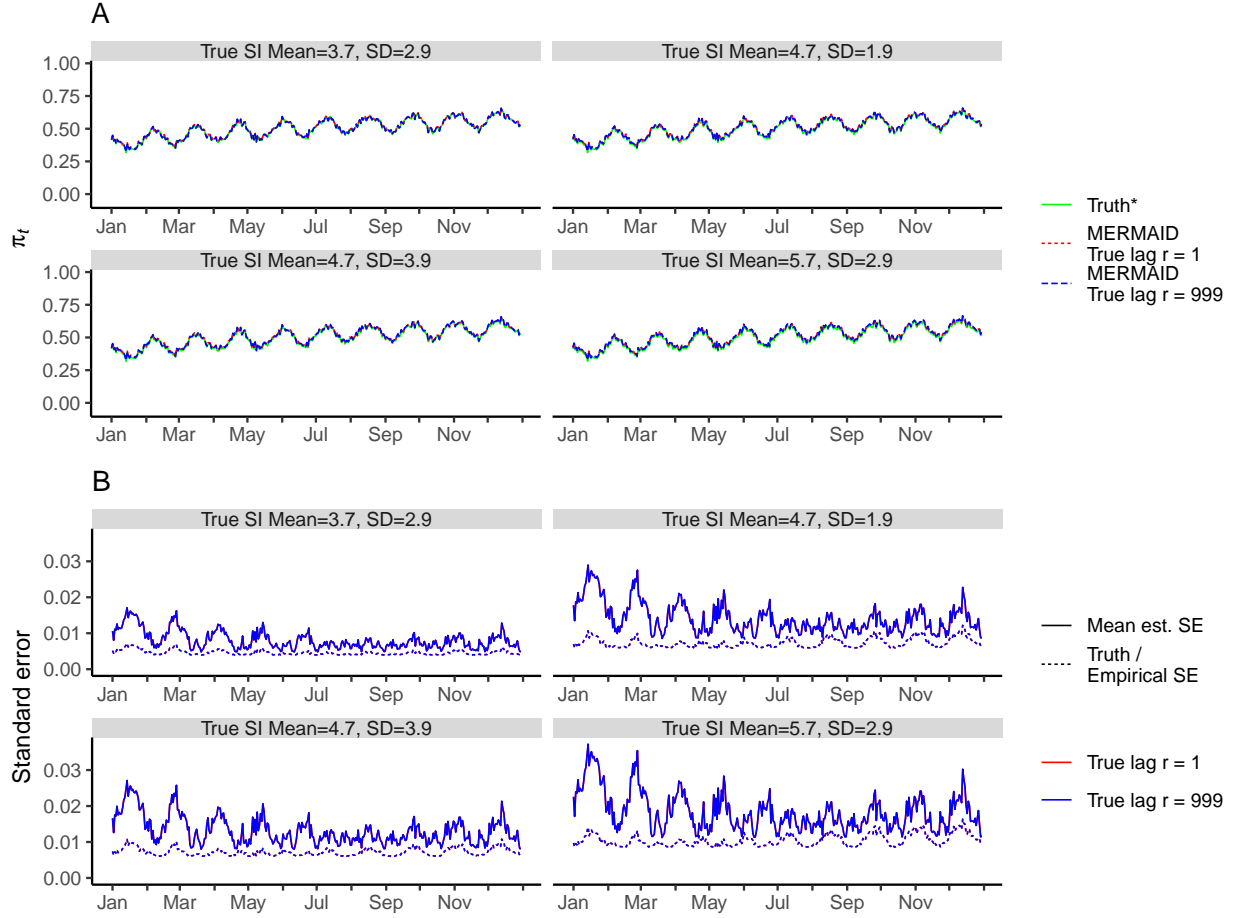
C Misspecified Simulations 1

In the following simulation settings, the fitted (assumed for the purpose of model fitting) infection-testing lag distribution is negative binomial with $\mu = 5$ and size $r = 5$ (truncated to 21 days), and the assumed serial interval distribution has mean 4.7 and standard deviation 2.9 (a discretized gamma density truncated to 60 days). The true data-generating infection-testing lag distribution (size parameter) and serial interval distribution (mean and standard deviation) are indicated in the figure. A total of 8 misspecified simulation settings are shown for each figure. The true (data-generating) R_{it} and π_{it} curves are held constant across these 8 settings.
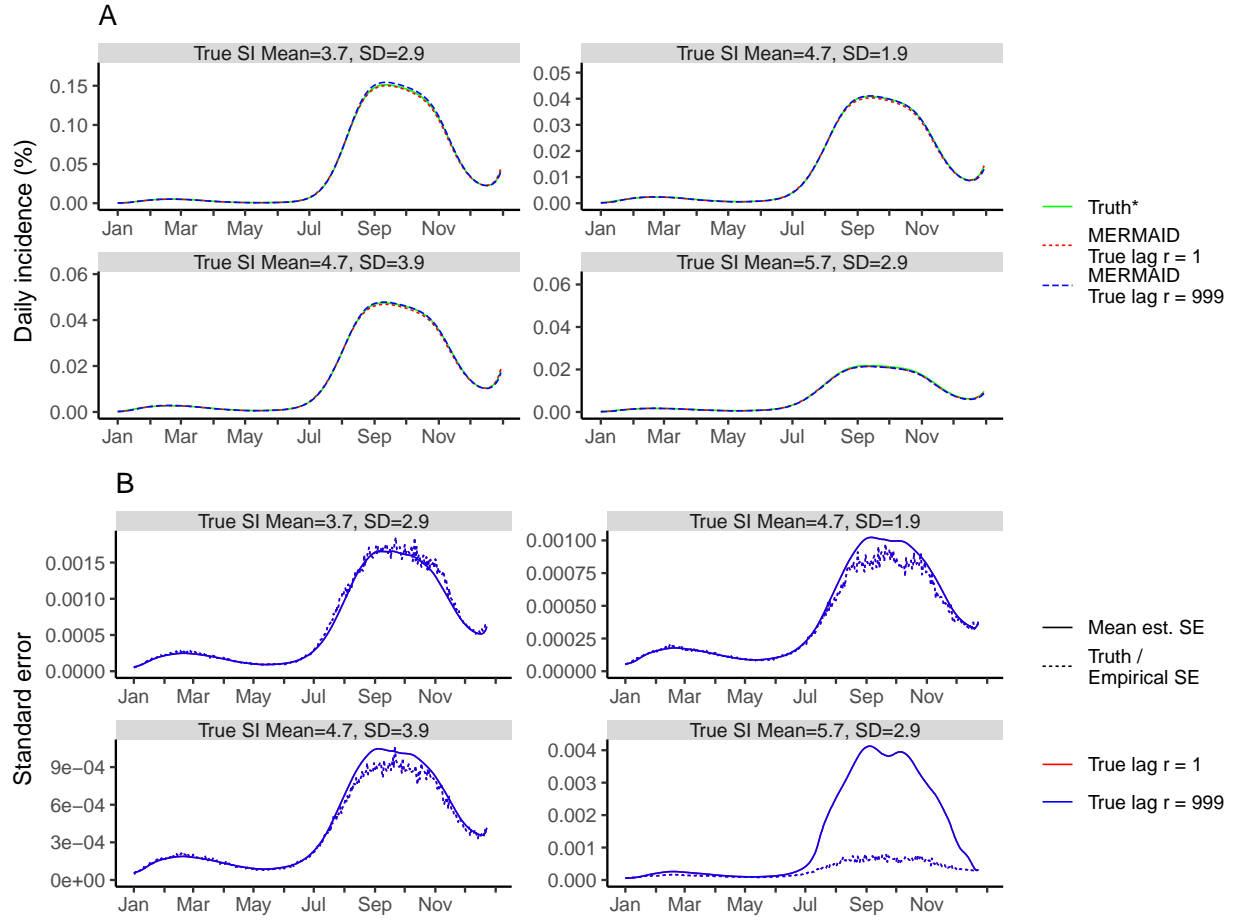
C.1 Effective reproductive numbers R_{it}
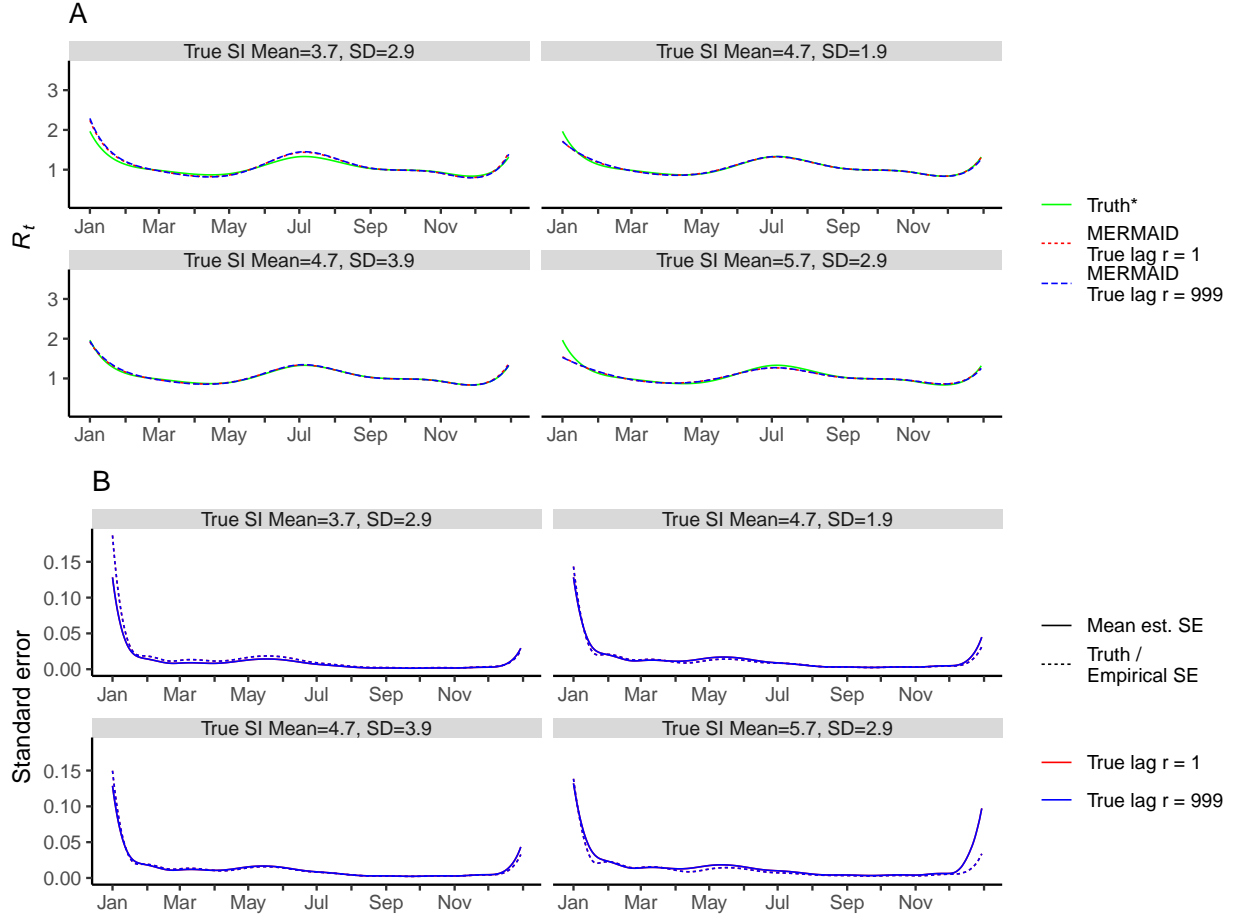
Supplementary Figure 9: Estimates of R_{it} in 8 simulation scenarios under a misspecified model. We vary the true (data-generating) serial interval (SI) parameters across 4 settings indicated in each plot facet, and the infection-reporting lag parameters across 2 settings indicated by line color (blue or red). **Panel A:** Shown are curves for true R_{it} (solid green), and $\hat{R}_{it}$ estimated using MERMAID (black). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).
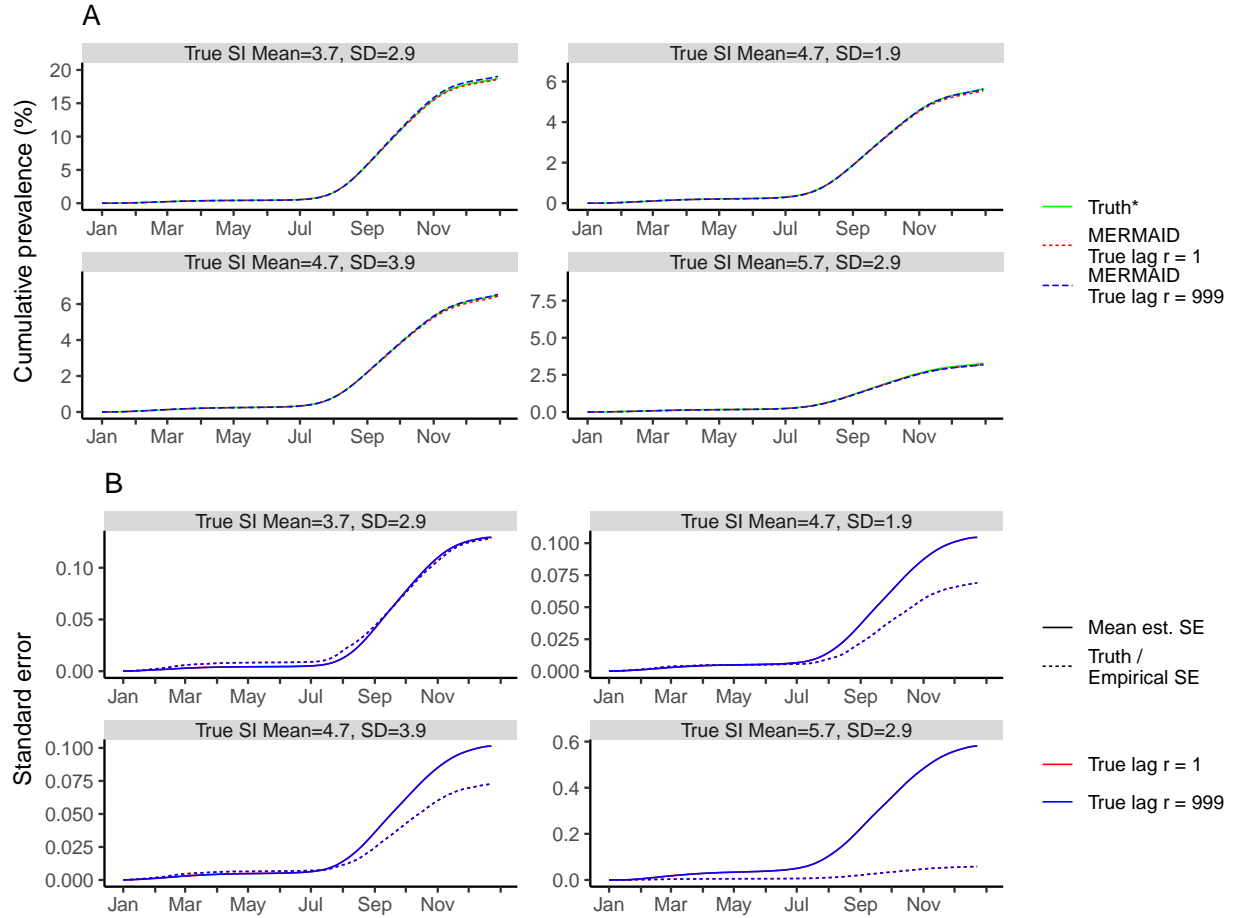
C.2 Ascertainment probabilities π_{it}

Supplementary Figure 10: Estimates of π_{it} in 8 simulation scenarios under a misspecified model. We vary the true (data-generating) serial interval (SI) parameters across 4 settings indicated in each plot facet, and the infection-reporting lag parameters across 2 settings indicated by line color (blue or red). **Panel A:** Shown are curves for true π_{it} (solid green), and $\hat{\pi}_{it}$ estimated using MERMAID (black). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates). Here, estimated standard errors are conservative due to the approximate correction for seroprevalence likelihood weights described in the main text.

C.3 Daily incidence

Supplementary Figure 11: Estimates of daily incidence in 8 simulation scenarios under a misspecified model. We vary the true (data-generating) serial interval (SI) parameters across 4 settings indicated in each plot facet, and the infection-reporting lag parameters across 2 settings indicated by line color (blue or red). **Panel A:** Shown are curves for true daily incidence (solid green), and estimates from MERMAID (black), GLM with raw reported case counts (magenta), or GLM with reconvoluted case counts (blue). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).

C.4   Cumulative prevalence

Supplementary Figure 12: Estimates of cumulative prevalence in 8 simulation scenarios under a misspecified model. We vary the true (data-generating) serial interval (SI) parameters across 4 settings indicated in each plot facet, and the infection-reporting lag parameters across 2 settings indicated by line color (blue or red). **Panel A:** Shown are curves for true cumulative prevalence (solid green), and estimates from MERMAID (black), GLM with raw reported case counts (magenta), or GLM with reconvoluted case counts (blue). **Panel B:** Shown are the root mean squared estimated standard error of MERMAID estimates, and the empirical mean standard error (the root-mean-square deviation of MERMAID estimates across replicates).