Supplemental Information for *Challenges in Estimating Time-Varying Epidemic Severity Rates from Aggregate Data*

Jeremy Goldwasser, Addison J. Hu, Alyssa Bilinski,
Daniel J. McDonald, Ryan J. Tibshirani

S1 Proofs and further analysis

The assumption of stationary delay distribution is not necessary for either bias expression. In the proofs in Appendix S1.1 and S1.2, the delay distributions π may simply be replaced with $\pi^{(t)}$.

S1.1 Proof of Proposition 1

Proposition 1 establishes the bias of the well-specified convolutional ratio. Here and henceforth, we abbreviate $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \{X_s\}_{s \le t}]$. Observe that

$$\begin{aligned}
\text{Bias}(\widehat{p}_t^{\pi}) &= \frac{\mathbb{E}_t[Y_t]}{\sum_{k=0}^{d} X_{t-k}\pi_k} - p_t \\
&= \frac{\sum_{k=0}^{d} X_{t-k}\pi_k p_{t-k}}{\sum_{k=0}^{d} X_{t-k}\pi_k} - \frac{p_t \sum_{k=0}^{d} X_{t-k}\pi_k}{\sum_{k=0}^{d} X_{t-k}\pi_k} \\
&= \sum_{k=0}^{d} \frac{X_{t-k}\pi_k}{\sum_{j=0}^{d} X_{t-j}\pi_j}(p_{t-k} - p_t).
\end{aligned}$$

S1.2 Proof of Proposition 2

Proposition 2 establishes the bias of the misspecified convolutional ratio, the lagged ratio being a special case. Observe that

$$\begin{aligned}
\operatorname{Bias}(\widehat{p}_t^{\gamma}) &= \frac{\mathbb{E}_t[Y_t]}{\sum_{k=0}^{d} X_{t-k}\gamma_k} - p_t \\
&= \frac{\sum_{k=0}^{d} X_{t-k}\pi_k p_{t-k}}{\sum_{k=0}^{d} X_{t-k}\gamma_k} - \frac{\sum_{k=0}^{d} X_{t-k}\gamma_k p_t}{\sum_{k=0}^{d} X_{t-k}\gamma_k} \\
&= \sum_{k=0}^{d} \frac{X_{t-k}}{\sum_{j=0}^{d} X_{t-j}\gamma_j}(\pi_k p_{t-k} - \gamma_k p_t) \\
&= \sum_{k=0}^{d} \frac{X_{t-k}}{\sum_{j=0}^{d} X_{t-j}\gamma_j}(\pi_k p_{t-k} - (\pi_k + (\gamma_k - \pi_k))p_t) \\
&= \frac{\sum_{j=0}^{d} X_{t-j}\pi_j}{\sum_{j=0}^{d} X_{t-j}\gamma_j} \sum_{k=0}^{d} \frac{X_{t-k}\pi_k}{\sum_{j=0}^{d} X_{t-j}\pi_j}(p_{t-k} - p_t) - p_t \sum_{k=0}^{d} \frac{X_{t-k}}{\sum_{j=0}^{d} X_{t-j}\gamma_j}(\gamma_k - \pi_k) \\
&= \frac{\sum_{j=0}^{d} X_{t-j}\pi_j}{\sum_{j=0}^{d} X_{t-j}\gamma_j} \operatorname{Bias}(\widehat{p}_t^{\pi}) + p_t \left(\frac{\sum_{k=0}^{d} X_{t-k}\pi_k}{\sum_{j=0}^{d} X_{t-j}\gamma_j} - 1 \right).
\end{aligned}$$

S1.3 Further analysis of well-specified bias

We present examples that further explain the bias for the well-specified convolutional ratio. These examples are considerably more contrived that the ones in Section 2.3. Nonetheless, their bias can be simplified to simple analytic formulas, isolating the three contributing factors.

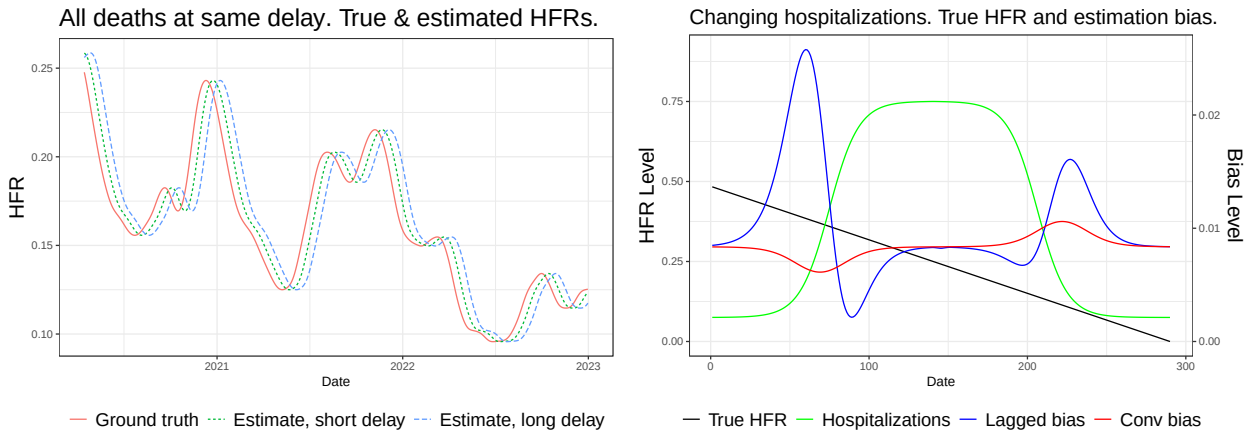
To elucidate the relationship between changing severity rates and bias, let us consider the case where all secondary events occur after precisely ℓ time points. The well-specified convolutional and lagged ratio estimators coincide: $\widehat{p}_t^{\gamma} = \widehat{p}_t^{\ell} = p_{t-\ell}$. The bias in this setting is simply the change in the true severity rate, $p_{t-\ell} - p_t$, and the ratio estimator is unbiased only if the severity rate is stationary. Otherwise the ratio will be 20% too low, for example, if the true severity rate was 20% lower ℓ time steps ago.

Figure S1a displays the results on the NHCS HFRs. In general, severity rates will be less similar to the present value p_t as we go further back in time. The bias $p_{t-\ell} - p_t$ tends to be larger when $\ell = 28$ versus $\ell = 14$. This supports the overarching idea that estimates with heavier-tailed delay distributions tend to have more bias.

Now to elucidate the relationship between primary incidence and bias, let us consider a delay distribution π which places half its mass at lag 0, and the other half at lag q. Then the well-specified bias has magnitude:

$$\left| \operatorname{Bias}(\widehat{p}_t^{\pi}) \right| = \frac{\frac{1}{2}\left| X_t(p_t - p_t) + X_{t-q}(p_{t-q} - p_t) \right|}{\frac{1}{2}(X_t + X_{t-q})} = \frac{|p_{t-q} - p_t|}{1 + X_t/X_{t-q}}.$$

In other words, the absolute bias is monotonically decreasing in X_t/X_{t-q}, the proportion



(a) All deaths after ℓ days. HFR ratios equivalent; plotting delays of $\ell = 14$ and 28 days.

(b) Changing primary incidence. Plotting bias of lagged and convolutional ratios.

Figure S1: Toy examples of biased severity rates.

change in primary incidence. Rising primary incidence ($X_t/X_{t-q} > 1$) yields less bias, while falling levels yield more.

Figure S1b displays this setting with $q = 10$. Rising hospitalizations are defined as $X = \sigma(s) * 9000 + 1000$, where σ is the sigmoid function and s takes 300 evenly spaced steps from -9 to 7. These quantities are subsequently reflected to express a decline. The true HFRs fall from 0.5 to 0 over the same number of even steps. Accordingly, the absolute bias of the convolutional ratio is $c_q \frac{X_{t-q}}{X_{t-q}+X_t}$, where $c_q \approx 0.0167$. Shown in red, it dips as hospitalizations rise, and rises as they fall.

The figure also plots with lagged ratio with $\ell = \frac{d}{2}$, the mean of the delay distribution. When daily hospitalizations are close to constant, the two estimators converge towards the same ratio. During periods of change, however, the lagged estimator has different bias. It first moves upwards — the opposite direction as the convolutional bias — with far greater magnitude. This can be explained by the ratio $A_t^\ell = \frac{X_{t-2\ell}+X_t}{2X_{t-\ell}}$ from Proposition 2. As hospitalizations begin to steeply rise, $X_{t-2\ell}$ and $X_{t-\ell}$ are similar, but $X_t > X_{t-\ell}$. Hence, $A_t^\ell > 1$, contributing positive bias to both the oracle and misspecification terms. As hospitalizations level out near the top, $A_t^\ell < 1$, hence the bias falling lower. The opposite pattern occurs as hospitalizations fall.

S2 Alternative data sources

S2.1 Retrospective deaths

JHU aggregated daily deaths in real time, aligned by the date they were reported. In contrast, the National Center for Health Statistics (NCHS) provided weekly totals of deaths aligned by occurrence, which were not available in real time. Intuitively, the delay which relates hospitalization to death occurrence should have a lighter tail in comparison to that relating hospitalization to death report. Therefore, since that heavier-tailed delays generally introduce

greater bias, we would expect the ratio estimates computed from JHU deaths to have greater bias than those from NCHS deaths. Figure S2 shows that this is indeed the case. NCHS data, which was only available in retrospect, leads to lagged HFR estimates with substantially less bias.

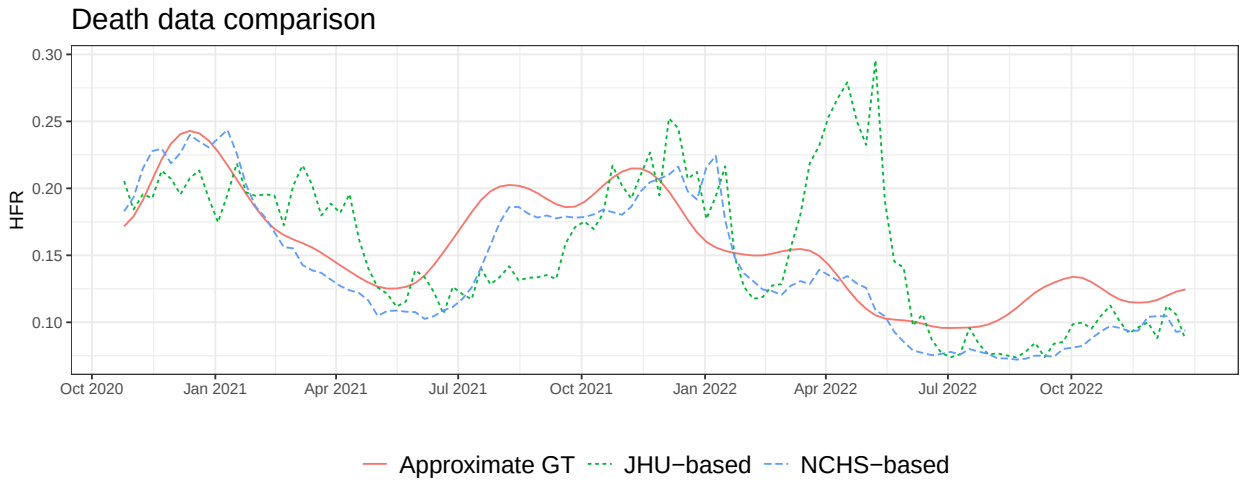


Figure S2: Comparing lagged ratios based on data from JHU versus NCHS.

S2.2 Alternative ground truth

We consider two alternative approaches to approximate the ground truth HFR curve, over the COVID-19 pandemic. The first is simply to use the lagged ratio based on NCHS data. As discussed above, this benefits from a lighter-tailed delay distribution than JHU data. Further, as we are looking to approximate ground truth in retrospect, we modify this estimator to be forward-looking: $\widehat{p}_t^{\ell} = Y_{t+\ell}/X_t$. The second approach we consider is to compute a single

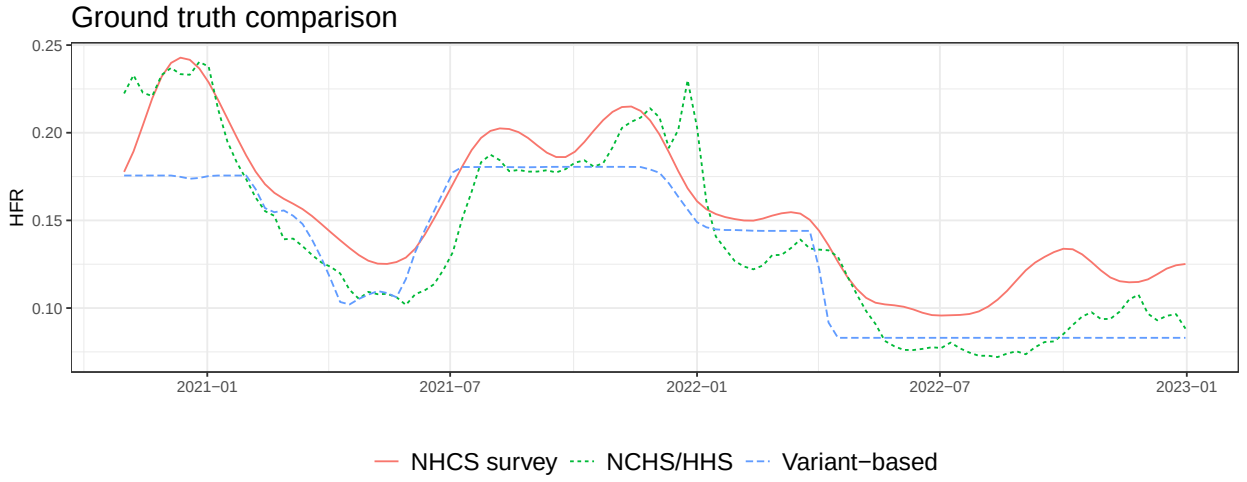


Figure S3: Comparing methods for approximating ground truth HFR.

HFR by dividing total deaths by total hospitalizations in each major variant period, and then create a smooth curve by mixing these per-variant HFRs by estimates of the proportion of variants in circulation, obtained from CoVariants.org. We only consider the four largest variants: the original strain, Alpha, Delta, and Omicron. However, because Omicron began with an enormous surge that quickly subsided, we split it into early and late periods.

Figure S3 displays these two alternative ground truth curves, alongside that obtained from NHCS. They have nontrivial differences in magnitude, but reassuringly, the three curves move more or less in conjunction. The retrospective NCHS ratio is still subject to statistical bias similar to equation (7); the variant-based HFR curve is flatter, as it does not account for other sources of potential variability in the underlying severity rate.

S3 Robustness checks

S3.1 Real-time versus finalized data

Recall, the results in Section 3.1 use hospitalization and death counts available in real time. To investigate the sensitivity of our findings, we recompute the lagged and convolutional ratios, this time using finalized counts. Figure S4 shows the real-time and finalized estimates track one another very closely. Therefore, the observed bias in Figure 3 cannot be attributed to real-time reporting quirks.

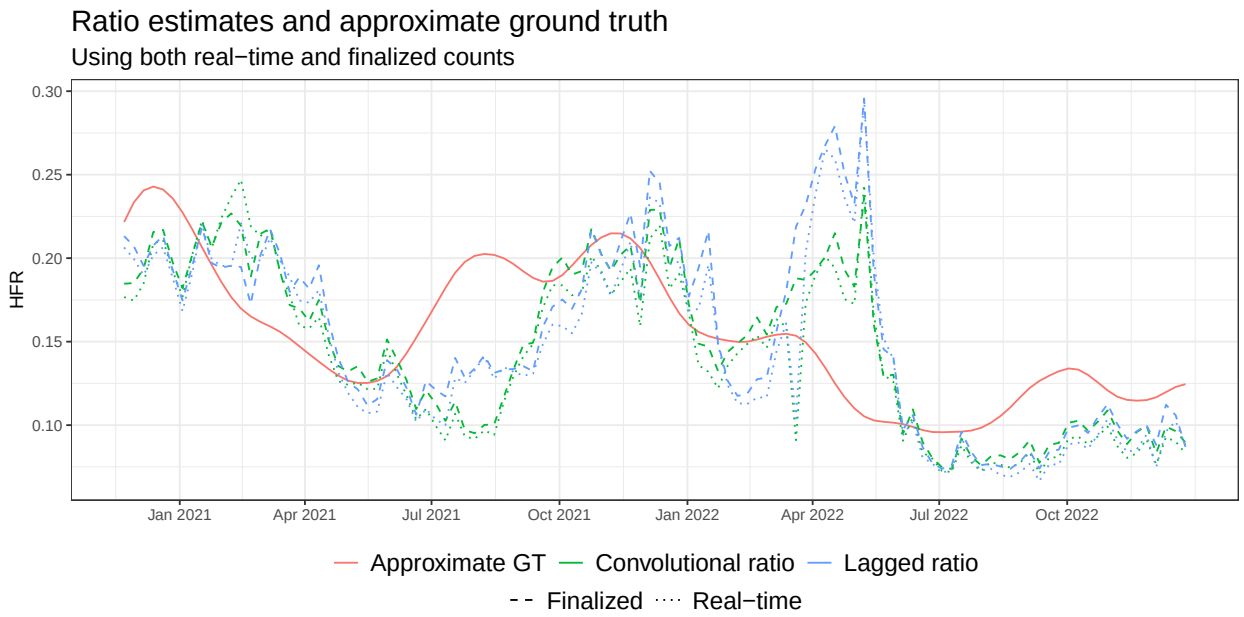


Figure S4: Comparing estimates based on real-time versus finalized counts.

S3.2 Hyperparameters

We evaluate the robustness of our findings against choices of hyperparameters. First, we analyze smoothed versions of the ratio estimators, where we smooth the numerator and

denominator separately:

$$\widehat{p}_t^{\ell,w} = \frac{\sum_{s=t-w+1}^{t} Y_s}{\sum_{s=t-w+1}^{t} X_{s-\ell}},$$
$$\widehat{p}_t^{\gamma,w} = \frac{\sum_{s=t-w+1}^{t} Y_s}{\sum_{s=t-w+1}^{t} \sum_{k=0}^{d} X_{s-\ell-k}\gamma_k}.$$

Figure S5 shows the results for varying window lengths $w > 0$. The results are very similar, indicating the bias does not disappear when smoothing over a longer history.

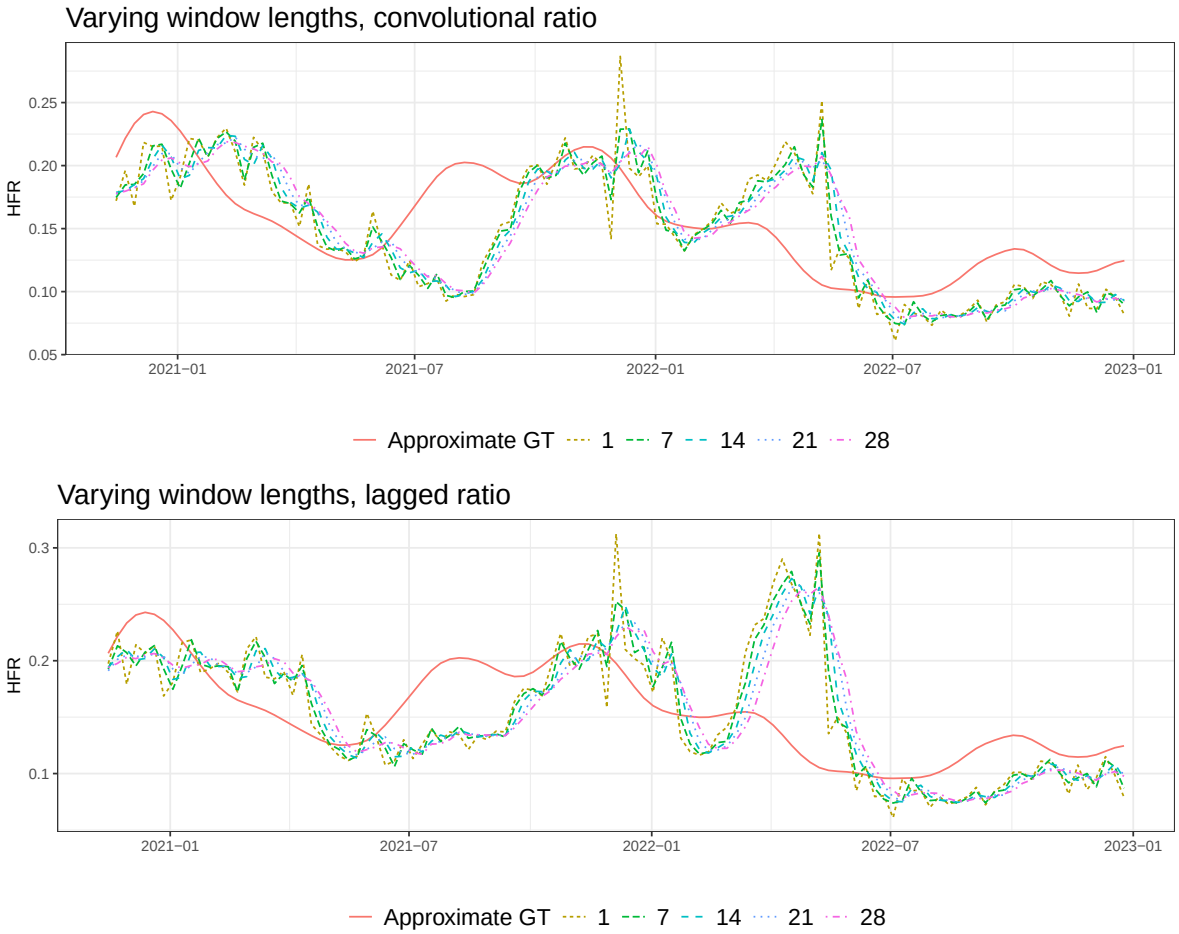


Figure S5: Comparing different choices of window length in post-smoothing.

We next examine the time-to-death hyperparameters: The lag for the lagged ratio and delay distribution for the convolutional ratio. Figure S6 displays lagged HFR estimates where the lag ℓ ranges from 2 to 5 weeks. Unlike the window size, changing this parameter leads to notably different behavior. Some choices of lag are better than others; a 28-day lag, for example, falls appropriately in winter 2021, and rises less slowly during Delta. However, all are biased to varying degrees, most notably the huge spurious surge in spring 2022.

Figure S7 compares the performance of the convolutional ratio across different choices of delay γ; we kept the discrete gamma shape for each, but varied the mean and standard

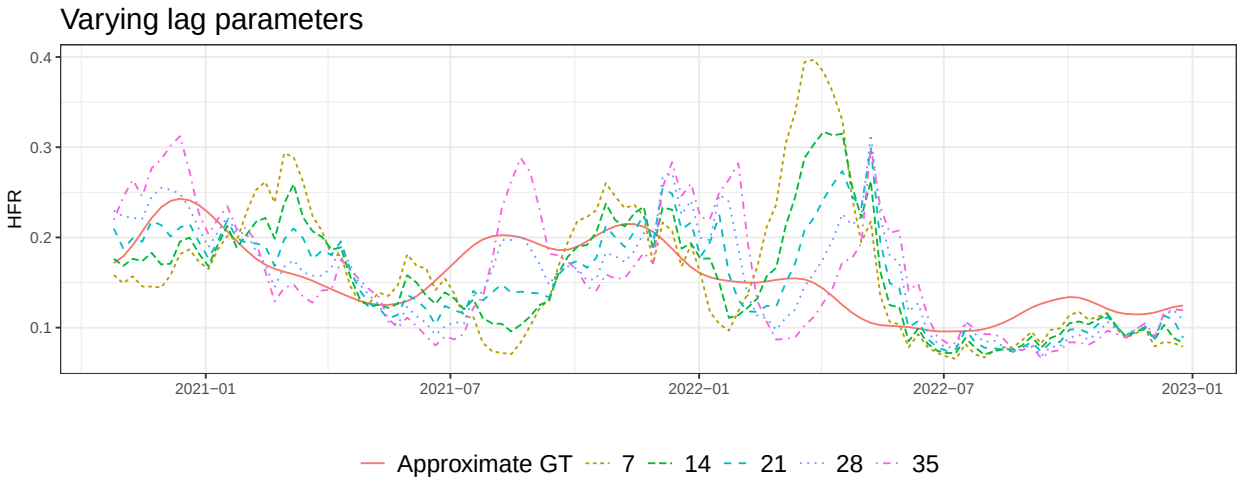


Figure S6: Comparing different choices of lag parameter in the lagged ratio.

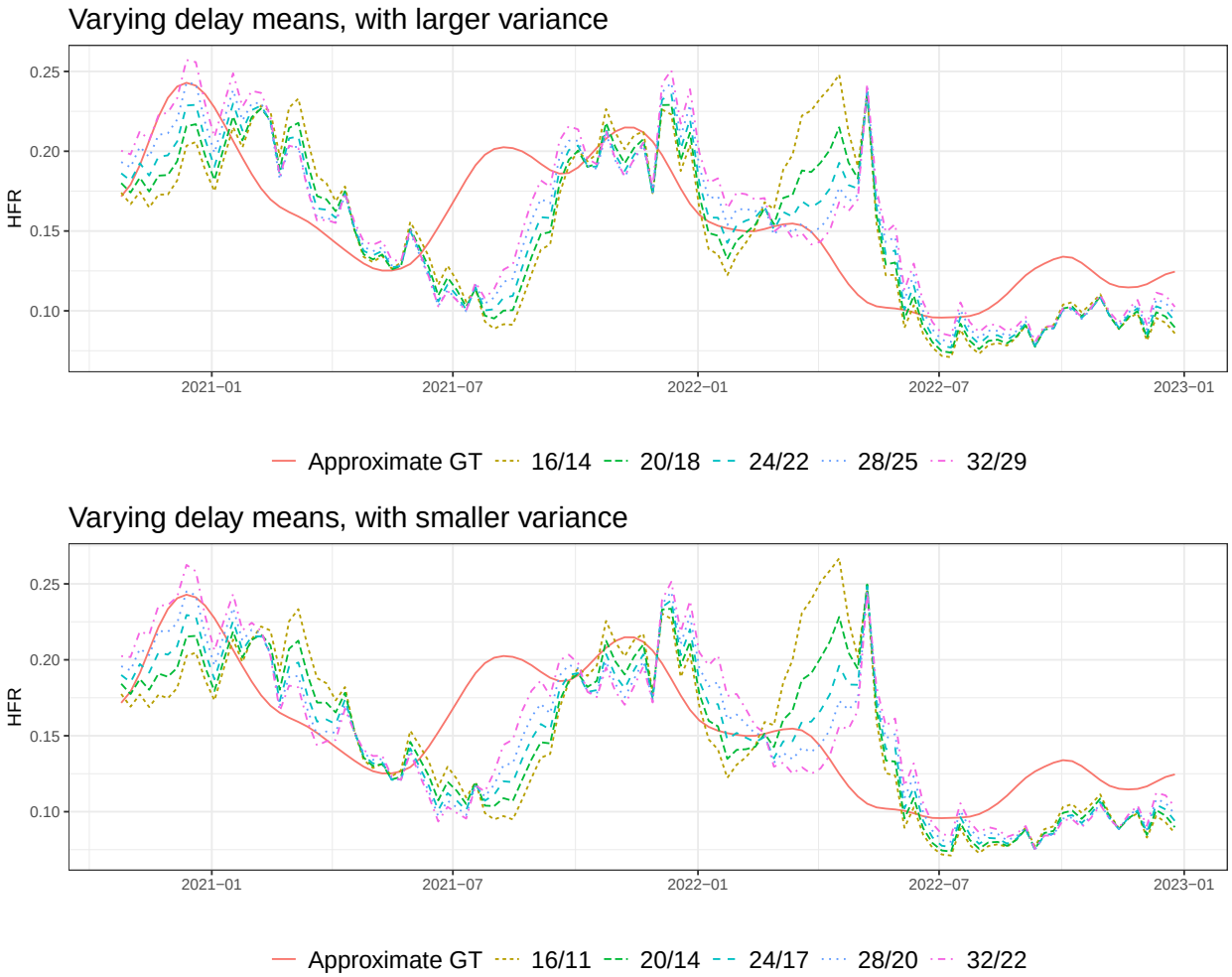


Figure S7: Comparing different choices of delay distributions in the convolutional ratio. The top panel shows gamma distributions whose standard deviation is 0.9 times the mean, versus 0.7 times the mean the bottom panel. The legend labels reflect the mean/standard deviation.

deviation. We investigated a standard deviation equal to 90% of the mean, and also a more compact delay with standard deviation equal to 70% of the mean. All resulting HFR estimates are significantly biased. Regardless of delay distribution, the ratios are negatively biased during the onset of Delta, and surge after the peak of Omicron. This indicates the bulk of the error is fundamental to the estimator, and cannot be attributed to model misspecification.

S3.3 State-level results

We repeat our analysis the six large US states, finding similar trends. The NHCS survey was conducted on a subset of hospitals meant to represent the US at large, so for the state-level approximate ground truth, we therefore use the forward-looking lagged ratio computed using NCHS deaths, as discussed above. Figure S8 compares this rough ground truth to the real-time convolutional and lagged ratios. For each state, we select the lag for the approximate ground truth curve to maximize cross-correlation between hospitalizations and NCHS deaths, and the lag in the real-time lagged ratio to maximize cross-correlation between hospitalizations and JHU deaths. We then use a discrete gamma distribution with mean equal to the latter lag, and standard deviation 90% of its mean, for the convolutional ratio.

Several states display similar biases to the US as whole (Figure 3). Estimates in California, Texas, and Florida are all slow to detect the uptick in HFR during Delta; they also spike during Omicron in California, and to a lesser extent Florida. Note these states are the ones with the largest cross-correlation optimal lags, an estimate of the average time-to-death. In contrast, New York, Pennsylvania, and Illinois have mean delays of at most 17. Their HFR estimates are still biased, but overall less so. This once again emphasizes the role of the delay in bias. The takeaway: fatality ratios are generally less trustworthy in states that take longer to report deaths.

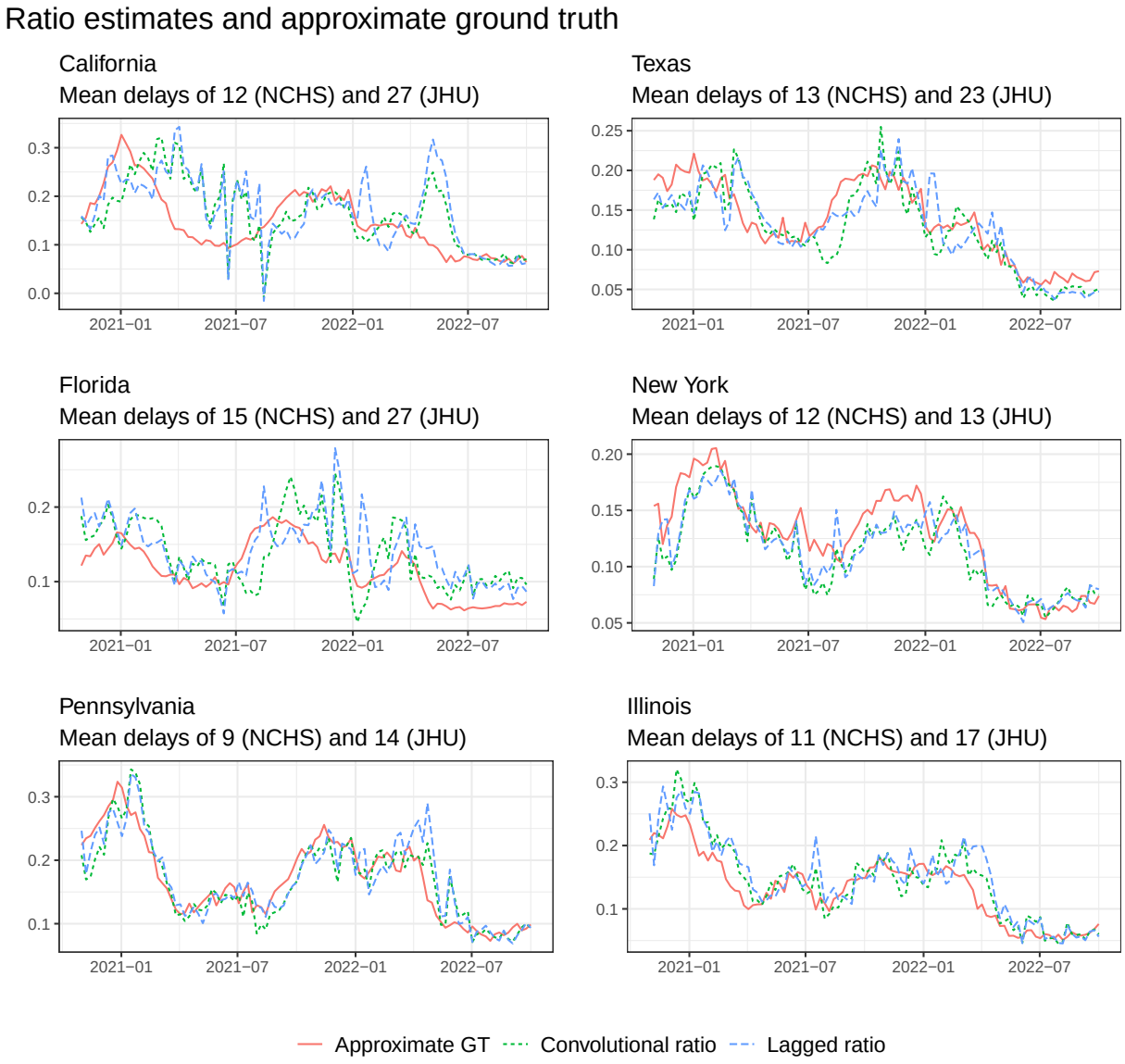


Figure S8: Comparing ratio estimates for individual US states.