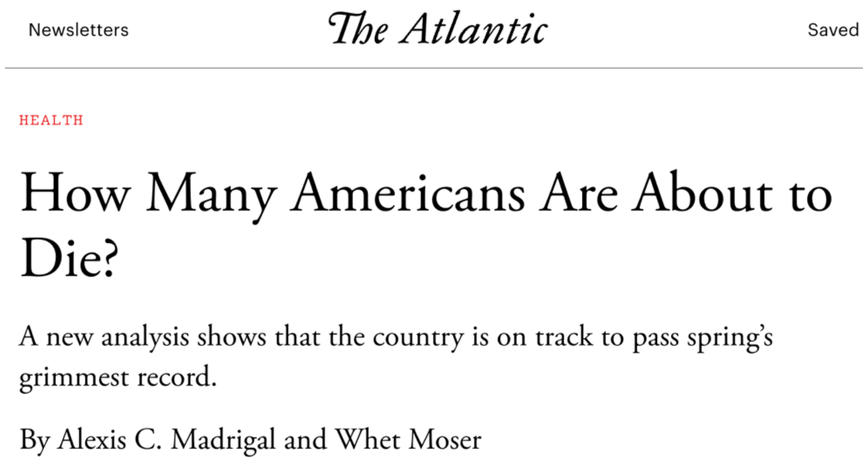
Estimating Epidemic Severity Rates

Jeremy Goldwasser

Time-varying severity rates in epidemiology

- Severity rates express the probability that a primary event at time t will result in serious secondary event, e.g.
  - Case-fatality rate (CFR)
  - Hospitalization-fatality rate (HFR)
- Time-varying or stationary?
  - Most academic work on estimating severity rates assumes stationarity over time.
  - Severity rates constantly change due to new variants, therapeutics, etc.
  - Epidemiologists at the CDC use time-varying rates to analyze new risks.



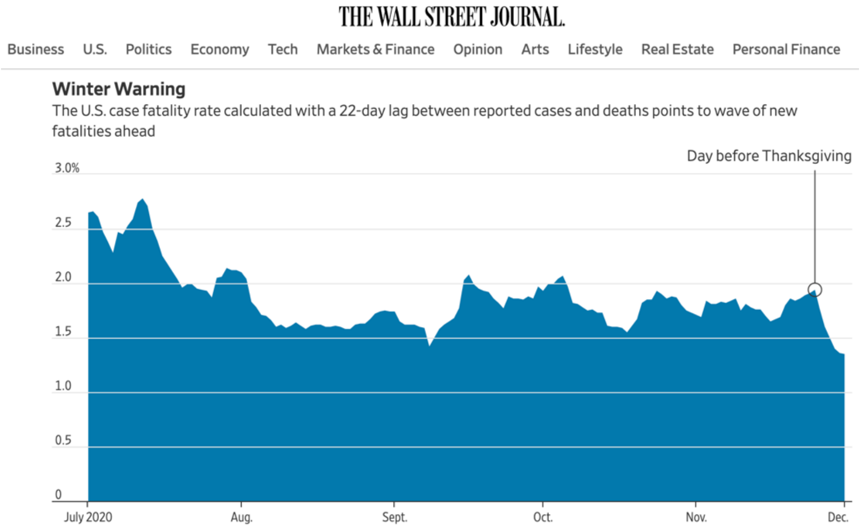
Newsletters The Atlantic Saved

HEALTH

How Many Americans Are About to Die?

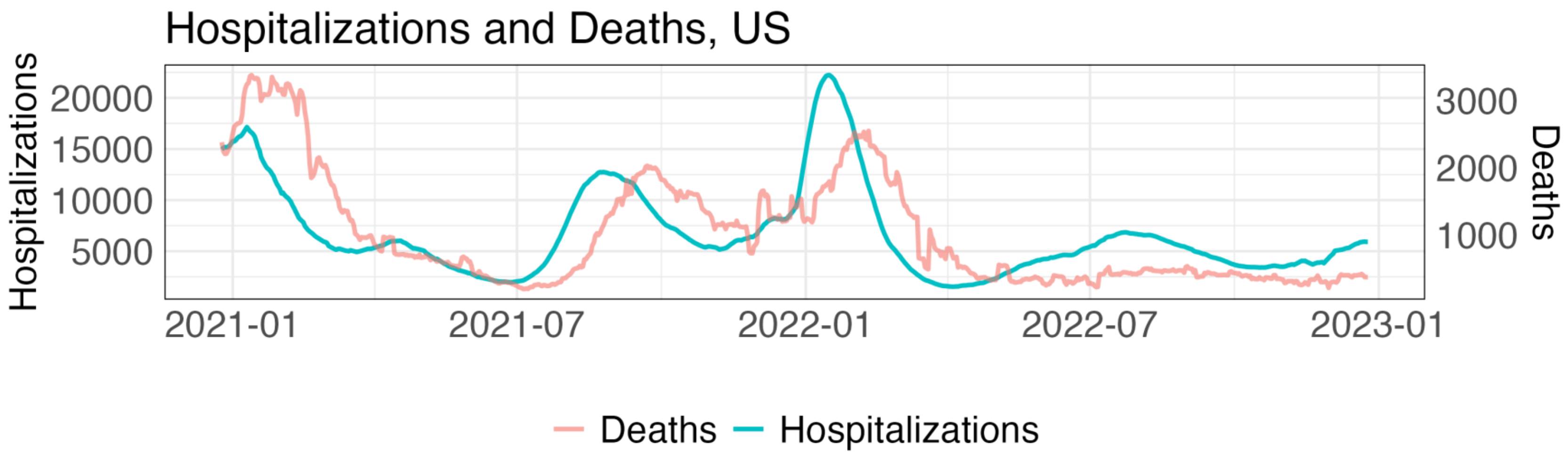
A new analysis shows that the country is on track to pass spring's grimmest record.

By Alexis C. Madrigal and Whet Moser



Often estimate severity from aggregate data

- Calculating severity rates is straightforward with a line list of patient outcomes.
  - CFR: Observe fraction of patients that tested positive at t who ultimately die.
- Maintaining such a line list may be unrealistic or impossible
  - In this case, severity rates must be estimated from aggregate count data.



Standard ratio estimators

- Most estimators for severity rates are simple ratios (“case fatality ratio”) between secondary events and at-risk primary events
- The standard time-varying approach is a lagged ratio of aggregate counts:

$$\widehat{\mathrm{CFR}}_t = \frac{\text{Deaths at } t}{\text{Cases at } t - \ell}$$

- A more principled generalization uses the delay distribution:

$$\widehat{\mathrm{CFR}}_t = \frac{\text{Deaths at } t}{\sum_k \{\text{Cases at } t - k\} \times \hat{\mathbb{P}}(\text{Death is at } k \text{ days})}$$

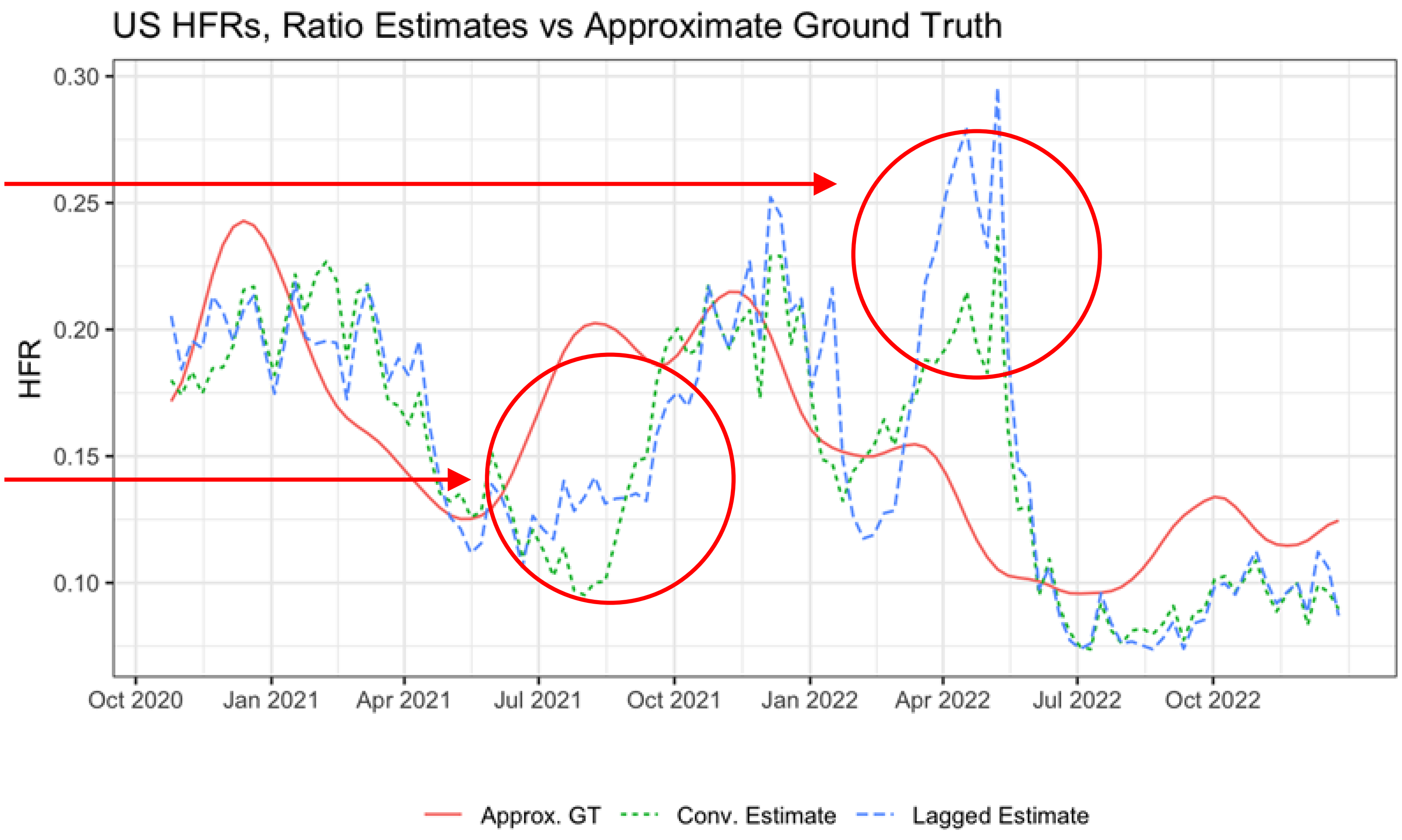
Our work: Understanding the bias of these ratios and proposing statistically sound alternatives.

Observed these ratios exhibit huge bias

Notable failures, HFR:

- Signaled enormous, nonexistent surge after Omicron peak – especially lagged ratio.
- Ignored higher risk as Delta took over

Findings robust across parameters, geography, etc.



Proposed solution: model the relationship between series

- Let $Y_t|X_{s \leq t}$ denote e.g. the number of deaths at time t given prior hospitalizations.

$$Y_t|X_{s\leq t} = \sum_{k=0}^{d} \sum_{i=1}^{x_{t-k}} \mathbf{1}\{i^{\text{th}} \text{ case at } t-k \text{ died at } t\}$$

- We identify this adheres to a *Poisson Binomial* distribution – a generalization of the binomial distribution where not all success probabilities are equal.
- While its PMF is intractable, it is well-approximated by a Gaussian with mean

$$\mu_t = \sum_{k=0}^{d} x_{t-k} \mathbb{P}(\text{die at } t \mid \text{hosp at } t-k) = \sum_{k=0}^{d} x_{t-k} \pi_k p_{t-k}$$

and variance

$$\sigma_t^2 = \sum_{k=0}^{d} x_{t-k} \pi_k p_{t-k} (1 - \pi_k p_{t-k}) \approx \mu_t.$$

Proposed solution: approximate MLE of probabilistic model

$$\hat{p}^{\text{MLE}}_{(t_0-d):T} = \text{argmax}_p \mathcal{L}(p) = \text{argmin}_p - \log \mathbb{P}(Y_t \; \forall t | X_{s\leq t} \; \forall t, \pi, p)$$

Correlation is negligible $\approx \text{argmin}_p \sum_{t=t_0}^{T} - \log \mathbb{P}(Y_t | X_{s\leq t}, \pi, p)$

Normal approximation at all t $\approx \text{argmin}_p \sum_{t=t_0}^{T} - \log \Phi(\frac{Y_t - \mu_t(p)}{\sigma_t(p)})$

Ignore variance term $\approx \text{argmin}_p \sum_{t=t_0}^{T} \frac{(Y_t - \mu_t(p))^2}{\sigma_t^2(p)}$

Plug-in variance $\approx \text{argmin}_p \sum_{t=t_0}^{T} \frac{1}{\hat{\mu}_t} (Y_t - \sum_{k=0}^{d} X_{t-k} \pi_k p_{t-k})^2$

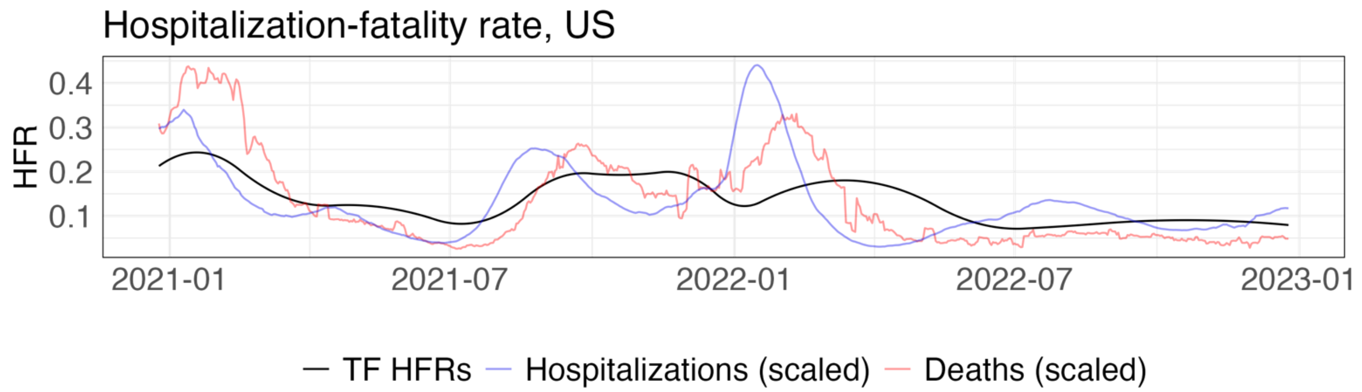
Plug-in delay distribution $\approx \text{argmin}_p \sum_{t=t_0}^{T} \frac{1}{\hat{\mu}_t} (Y_t - \sum_{k=0}^{d} X_{t-k} \gamma_k p_{t-k})^2.$

Proposed solution: learn severity rates with smoothed MLE

- To find a smooth solution for this overparameterized problem, we maximize the likelihood subject to a *trend filtering* penalty.

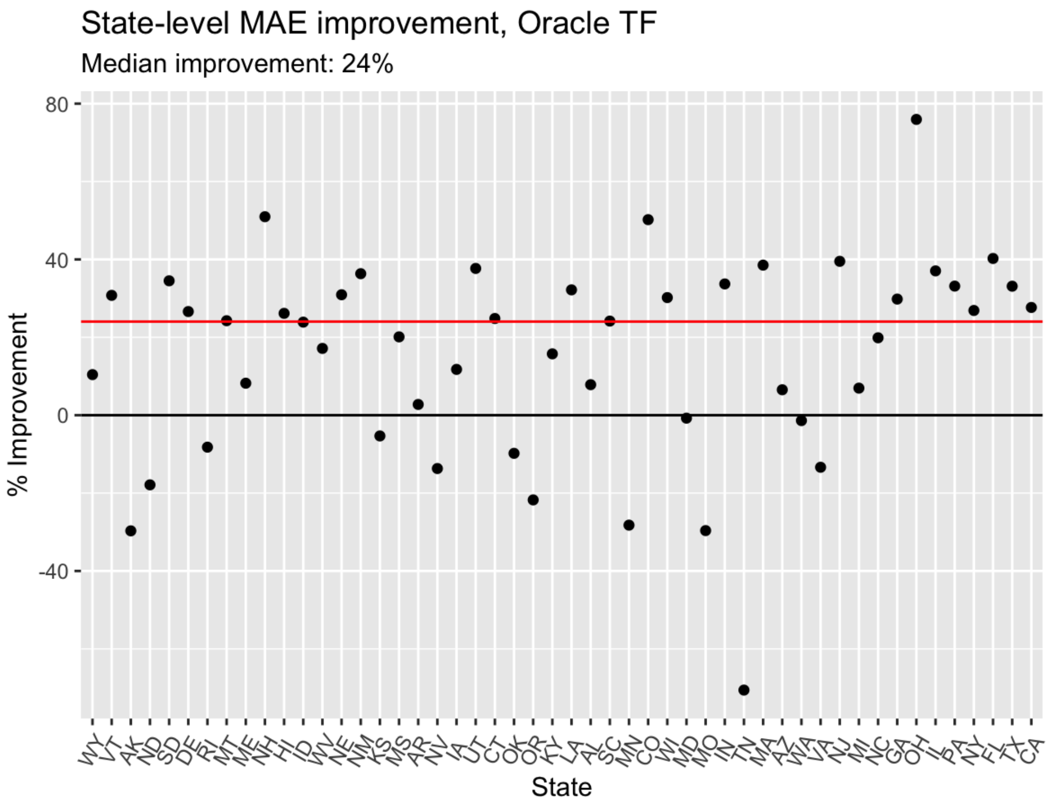
$$\hat{p}^{\mathrm{TF}} = \mathrm{argmin}_{p \geq 0} \sum_{t=t_0}^{T} \frac{1}{\hat{\mu}_t}(Y_t - \sum_{j=0}^{d} X_{t-j}\gamma_j p_{t-j})^2 + \lambda \| D^{(k+1)} p \|_1$$

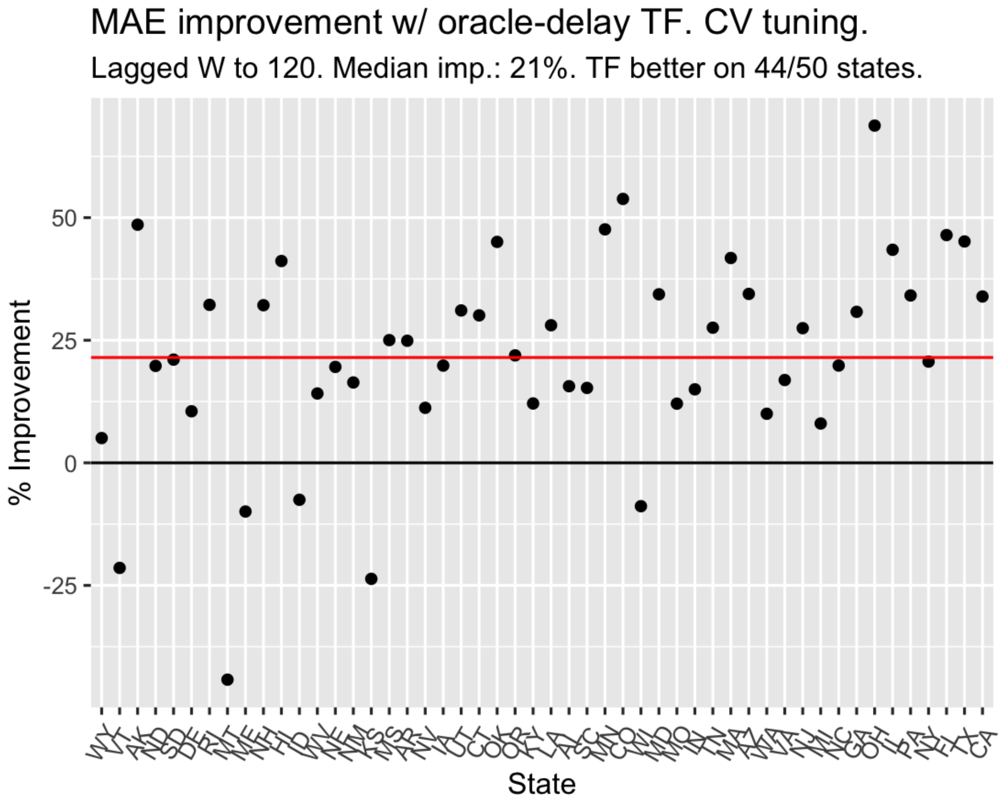
- The difference matrix $D^{(k+1)}$ contains finite differencing operations of order *k+1*. The L1 penalty encourages *p* to have sparse *k+1*th discrete derivatives, so solutions are piecewise polynomials of order *k*.
  - Trend filtering is more locally adaptive than smoothing splines.



Trend filtering estimator outperforms lagged estimator

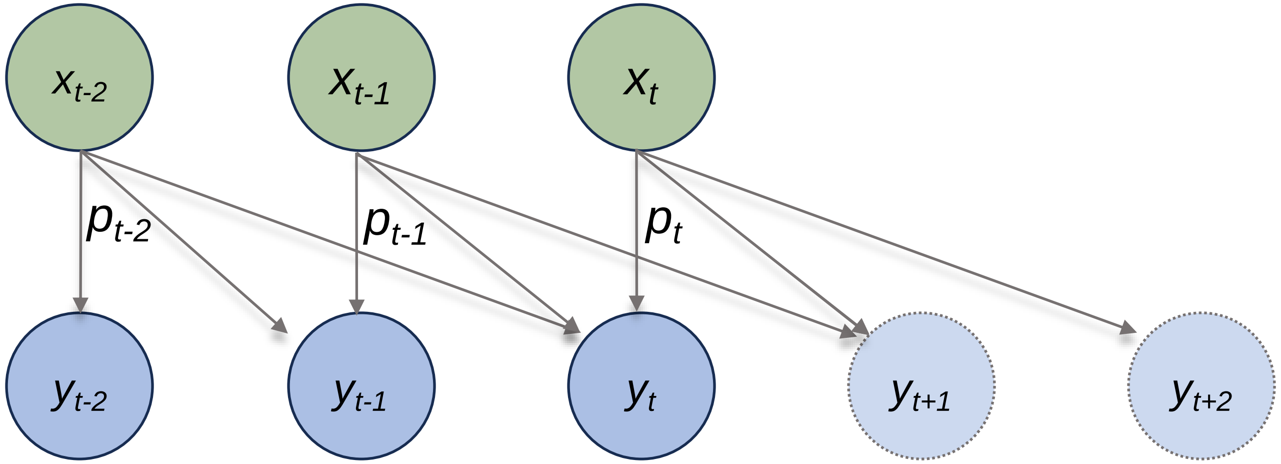
- State-level deaths simulated from overdispersed probabilistic model.
- On average, trend filtering **lowers MAE by >20%** over the lagged estimator – with both cross validation and oracle tuning.





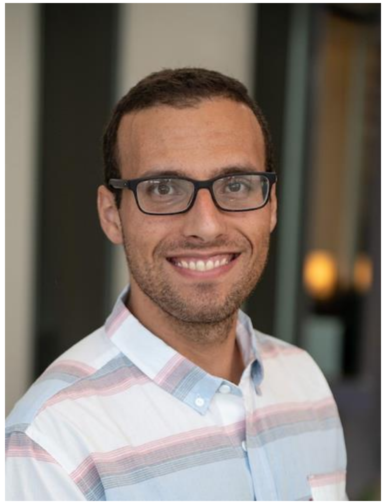
Ongoing: Adapt trend filtering for real-time setting

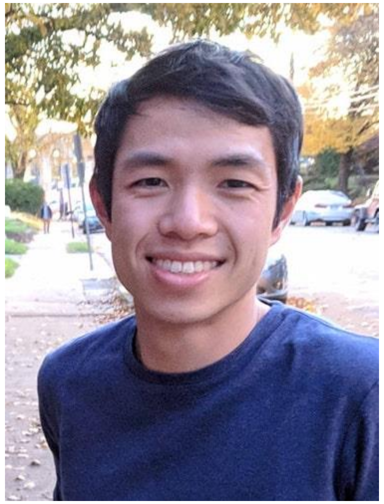
- Requires extra regularization to mitigate tail variability, since most recent severity rates used for fewer observed predictions.



- Jahja et al. (2022) used natural trend filtering & tapered smoothing for similar deconvolution problem.
- Also aim to quantify uncertainty of severity estimates and compare to convolutional ratio.

Collaborators





Thanks for your attention!