Supporting Information for

Adaptive metrics for an evolving pandemic
A dynamic approach to area-level COVID-19 risk designations

Alyssa Bilinski, Joshua A. Salomon, and Laura Hatfield

Corresponding Author: Alyssa Bilinski.
E-mail: alyssa_bilinski@brown.edu

This PDF file includes:

Supporting text
Figs. S1 to S12

Supporting Information Text

A. Population-Weighted Performance. Let S_i be the population of location i. The weight for location i is $s_i = \frac{S_i}{\sum_{i=1}^{N} S_i}$. To obtain population-weighted estimates of our accuracy measure, we estimate:

$$\delta_{wt} = 1 - \sum_i s_i \left(p_{FP} w_P + p_{FN} w_N\right)$$

B. Optimal Threshold Selection. We first specify a weight wt such that we consider false positives to be a factor of wt as costly as false negatives. For example, if $wt = 3$, we consider 3 false positives as costly as 1 false negative; 2 false positives would be less costly than 1 false negative. Denote the probability that an outcome of interest occurs given observed indicators for an observation, $q_{i,w+3} = Pr(Y_{i,w+3} = 1|X_i)$. We want to predict that the outcome will occur if, in expectation, this will decrease net costs. If an observation with probability $q_{i,w+3}$ is classified with a prediction of 0, this has probability $q_{i,w+3}$ of being a false negative. If it is classified with a prediction of 1, there is a probability $1 - q_{i,w+3}$ of a false positive. We therefore should classify with a prediction of 1 if:

$$\begin{aligned} \text{Expected cost of FP} &\leq \text{Expected cost of FN} \\ (1 - q_{i,t+3}) &\leq q_{i,t+3} wt \\ q_{i,t+3} &\geq \frac{1}{1 + wt} \end{aligned}$$

C. Simulations. To conduct simulations, we generate data that assumes a logistic relationship between the probability of a high outcome ($Pr(Y_{i,w+3})$) and a synthetic hospitalization indicator ($X_{H,i,w}$):

$$logit\left(Pr(Y_{i,w+3} = 1|\, X_{H,i,w}\right) = \beta_0 + \beta_1 X_{H,i,w} + \epsilon_{iw}, \quad [1]$$

where ϵ_{iw} are $i.i.d$ draws from a logistic distribution with mean 0 and scale parameter σ. We then draw $Y_{i,w+3}$ from a binomial distribution with the corresponding probability.

We vary simulations across 3 main dimensions:

1. **Indicator prevalence:** We first use empirical hospitalization data for simulations, drawing synthetic outcomes according to 1. To build intuition, we then use two stylized scenarios, one in which prevalence is constant over quarters and one in which waves are even more pronounced.
   (a) Empirical: We use true state-level hospitalization data from Q3 2021 through Q4 2022.
   (b) Constant: We draw $X_{H,i,w}$ from a $Unif(2, 20)$ distribution for state-times from Q3 2021 through Q4 2022.
   (c) Sharp waves: We alternate each quarter between drawing hospitalizations from a $N(5, 1)$ distribution and a $N(15, 1)$ distribution for each state-time from Q3 2021 through Q4 2022.
2. **Relationship between inputs and outputs:** The optimal cutoff for a metric with neutral weighting is $-\beta_0/\beta_1$. We vary this as displayed in Figure S11:
   (a) Constant: 10 hospitalizations per 100,000 population
   (b) Linear increase: linearly increasing from 5 to 15 hospitalizations per 100,000 over the study period
   (c) Logistic increase: increasing from 5 to 15 hospitalizations per 100,000 over the study period per a logistic model with a sharp increase at week 25
   (d) Non-monotonic: optimal cutoff increases and then decreases

For illustration, we set $\beta_1 = 3, \sigma = 1$ and use the first quarter (synthetic Q3 2021) as training data. For each scenario, we simulate 50 draws. As in the main text, we select the best-performing static metric during training data and compare performance in terms of predictive accuracy to adaptive metrics, averaging over draws. Results are displayed in Figure S12.

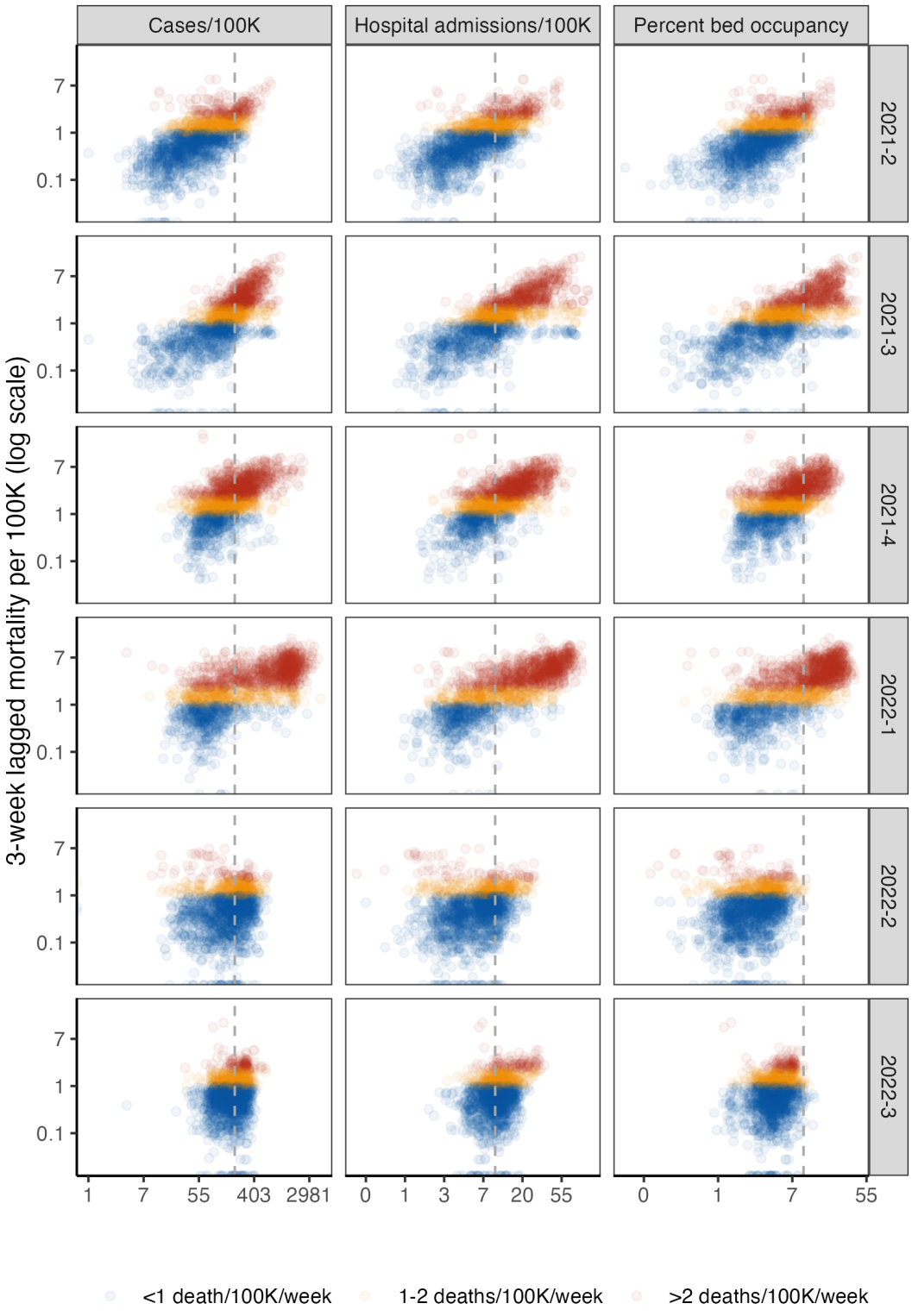


Fig. S1. County-level lagged mortality vs. indicator levels by quarter. Columns indicate different indicators (weekly cases per 100,000 population, new hospital admissions per 100,000, and percentage of inpatient beds occupied by COVID-19 patients), and rows indicate quarters. The x-axis displays indicator values on a log scale and y-axis displays 3-week ahead mortality per 100,000 population on a log scale. Each point on the scatterplot is a county-week. Colors show mortality outcome level. The vertical gray dotted lines indicate thresholds from CDC Community Levels for each indicator (≥ 200 cases/100K/week and ≥ 10 new admissions/100K/week or $\geq 10\%$ COVID-19 bed occupancy.)

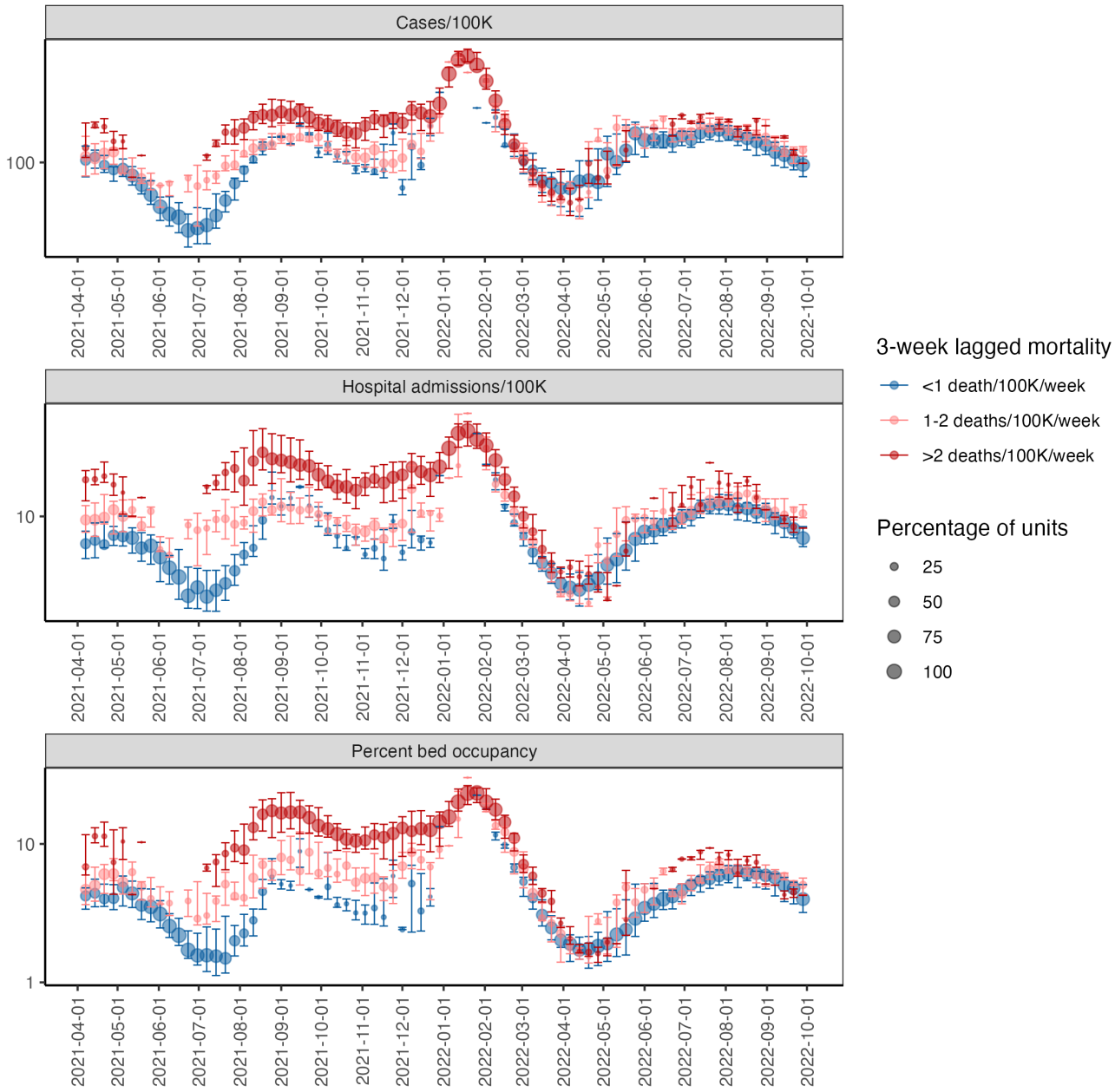


Fig. S2. Indicators by lagged mortality (state). Indicators vary across rows. The x-axis displays time and the y-axis displays the median (point) and interquartile range (bars) of each indicator by future mortality status.

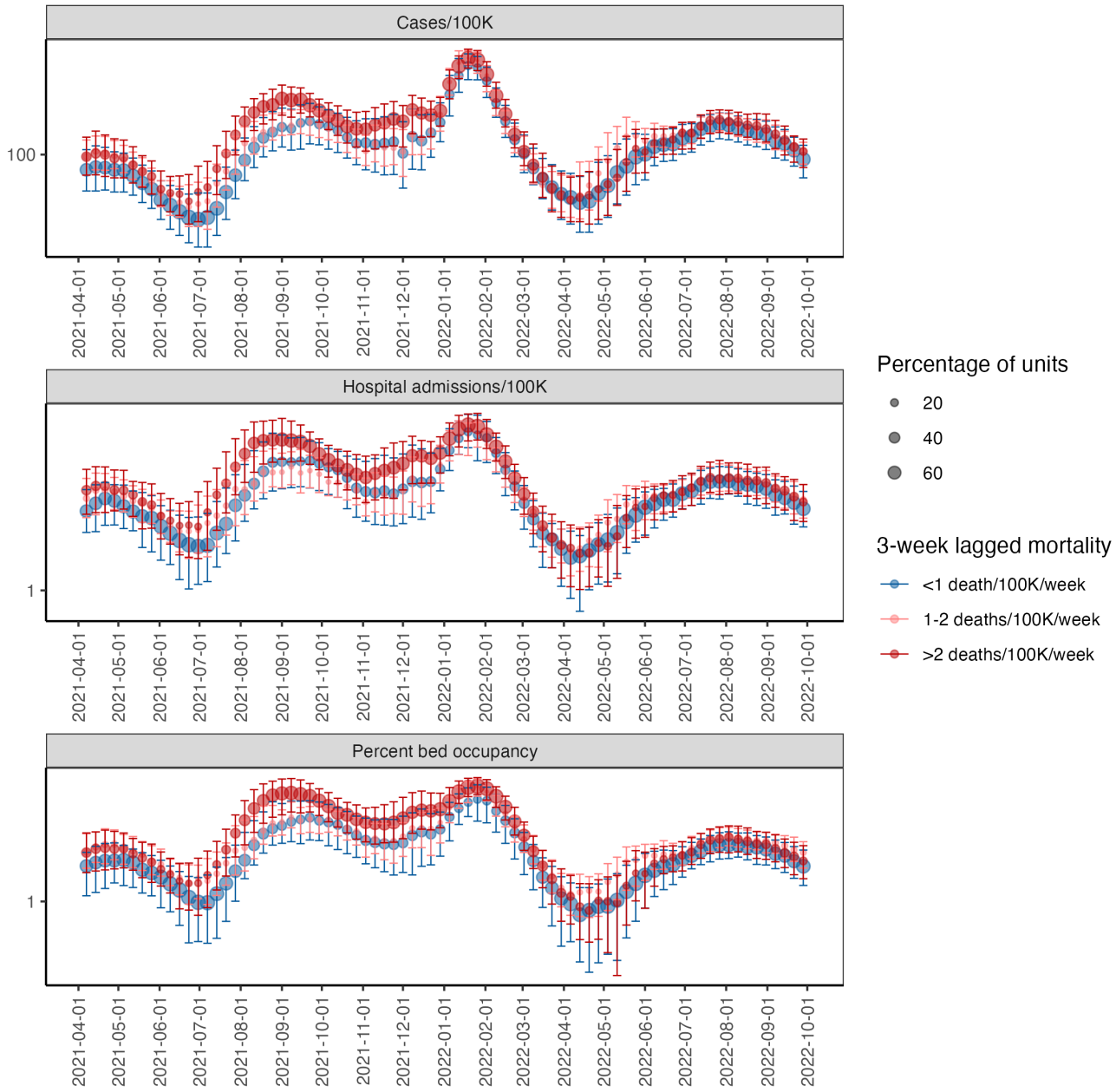
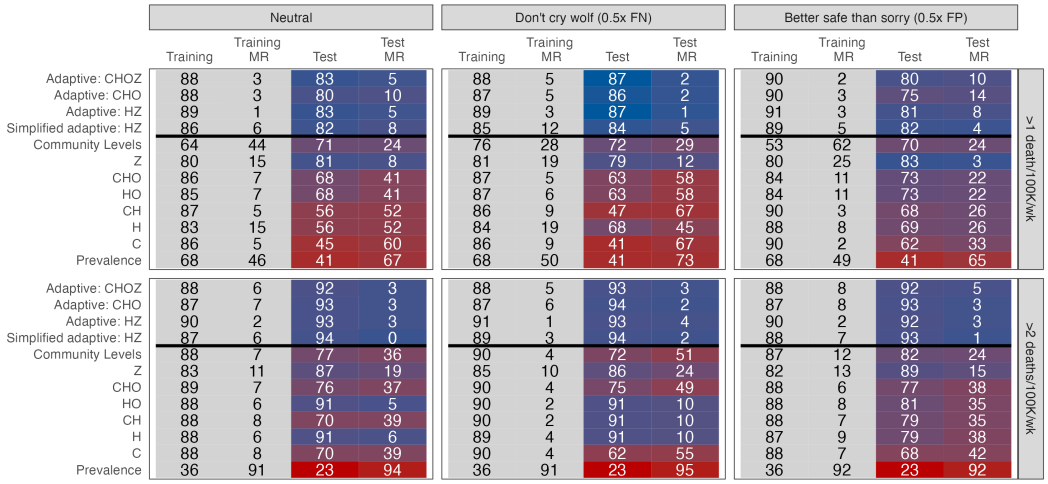
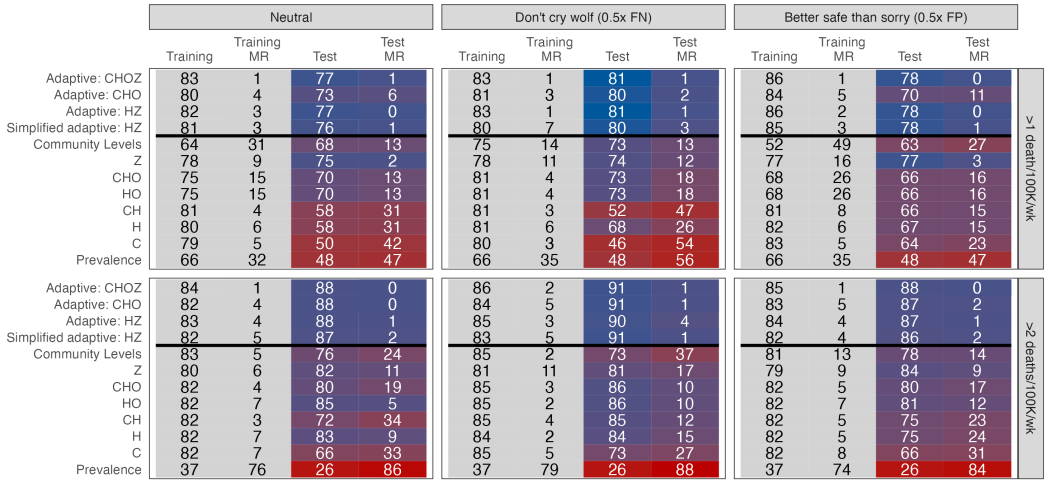


Fig. S3. Indicators by future mortality (county). Indicators vary across rows. The x-axis displays time and the y-axis displays the median (point) and interquartile range (bars) of each indicator by future mortality status.

States



HSAs



Counties

Fig. S4. Head-to-head comparison results. The top plots display results from state-level analyses and the bottom plots display results from county-level analyses, both weighted for population. Metrics are displayed on the left, with training data from Q2-Q4 2021 and test data from Q1-Q3 2022. Cells report weighted accuracy and maximum regret (MR) over training and test periods. Rows vary outcomes, and columns vary preferences for false positive versus false negatives, with "neutral" corresponding to unweighted accuracy. Prevalence indicates the proportion of high location-weeks in a given time period. Weighted accuracy by quarter, including for intensive care usage, is presented in Figures S5-S7.

States

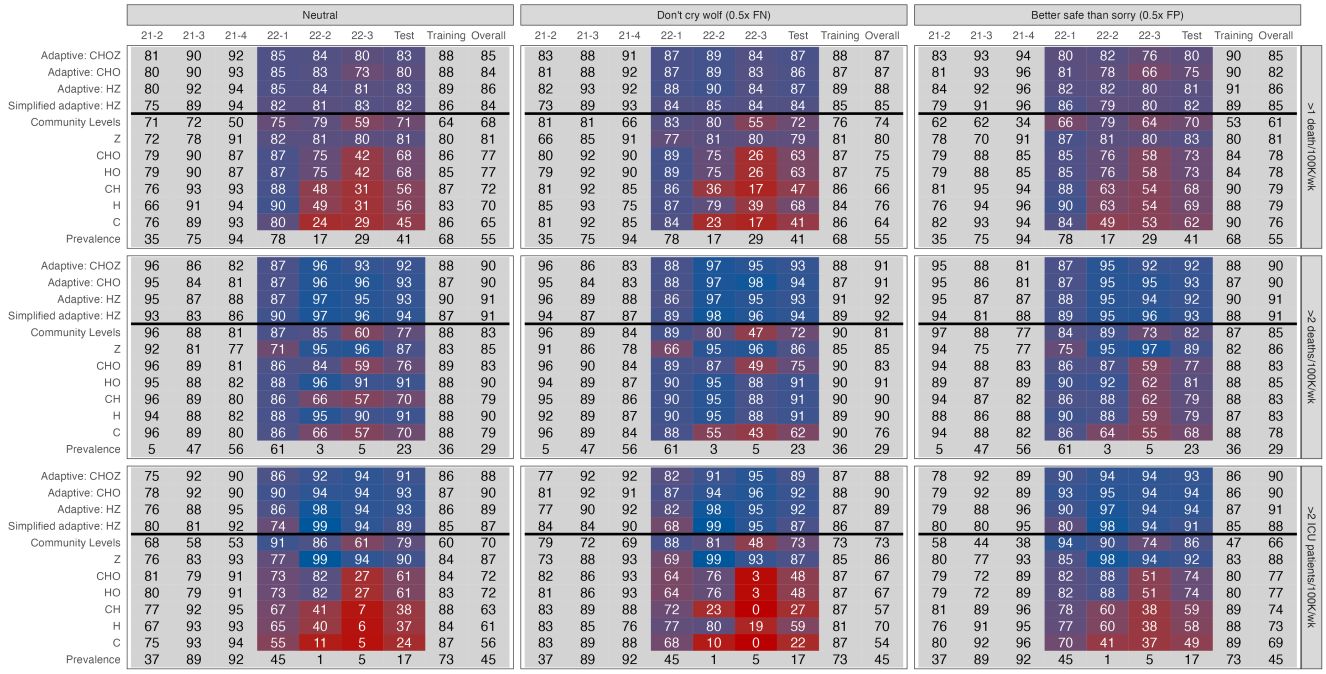


Fig. S5. State-level results by quarter. Metrics are displayed on the left, with training data from Q3-Q4 2021 and test data from Q1-Q3 2022. Cells report weighted accuracy. Preferences for false positive versus false negatives varying across columns (with "neutral" corresponding to unweighted accuracy) and outcomes across rows. Prevalence indicates the proportion of high location weeks in a given quarter.

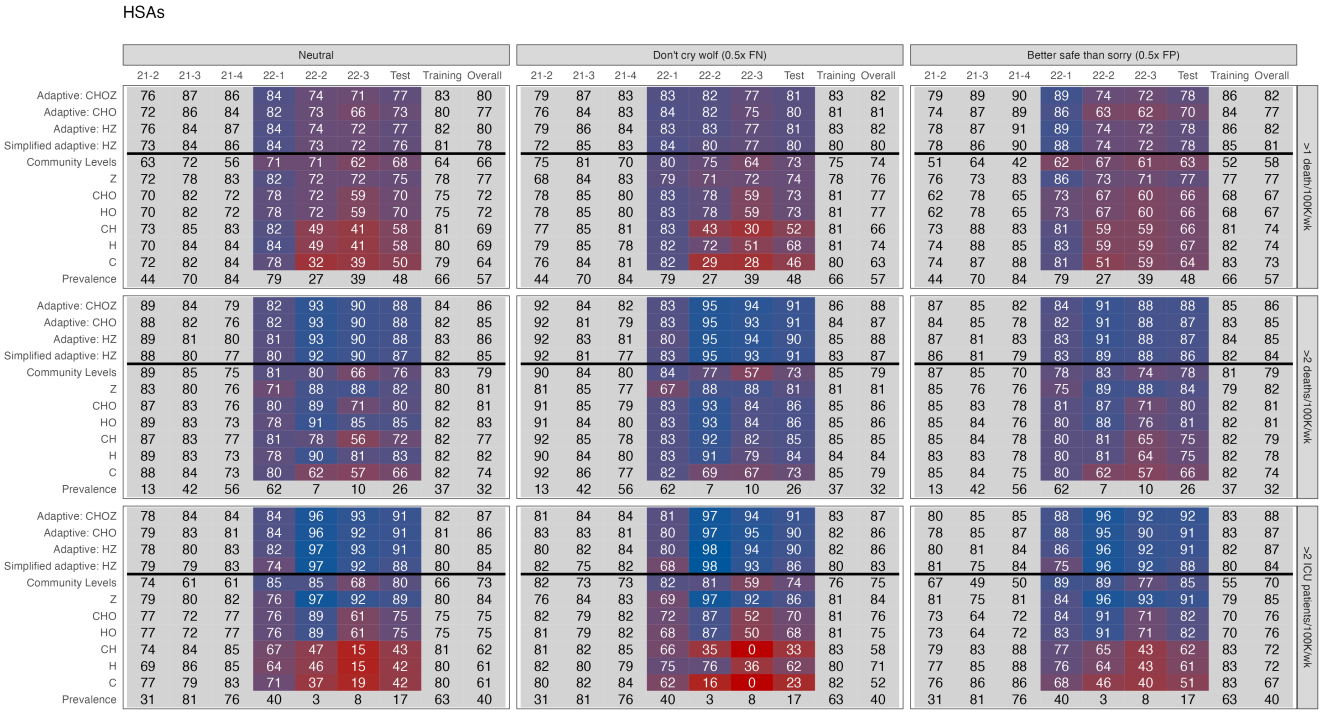


Fig. S6. HSA-level results by quarter. Metrics are displayed on the left, with training data from Q3-Q4 2021 and test data from Q1-Q3 2022. Cells report weighted accuracy. Preferences for false positive versus false negatives varying across columns (with "neutral" corresponding to unweighted accuracy) and outcomes across rows. Prevalence indicates the proportion of high location weeks in a given quarter.

Counties

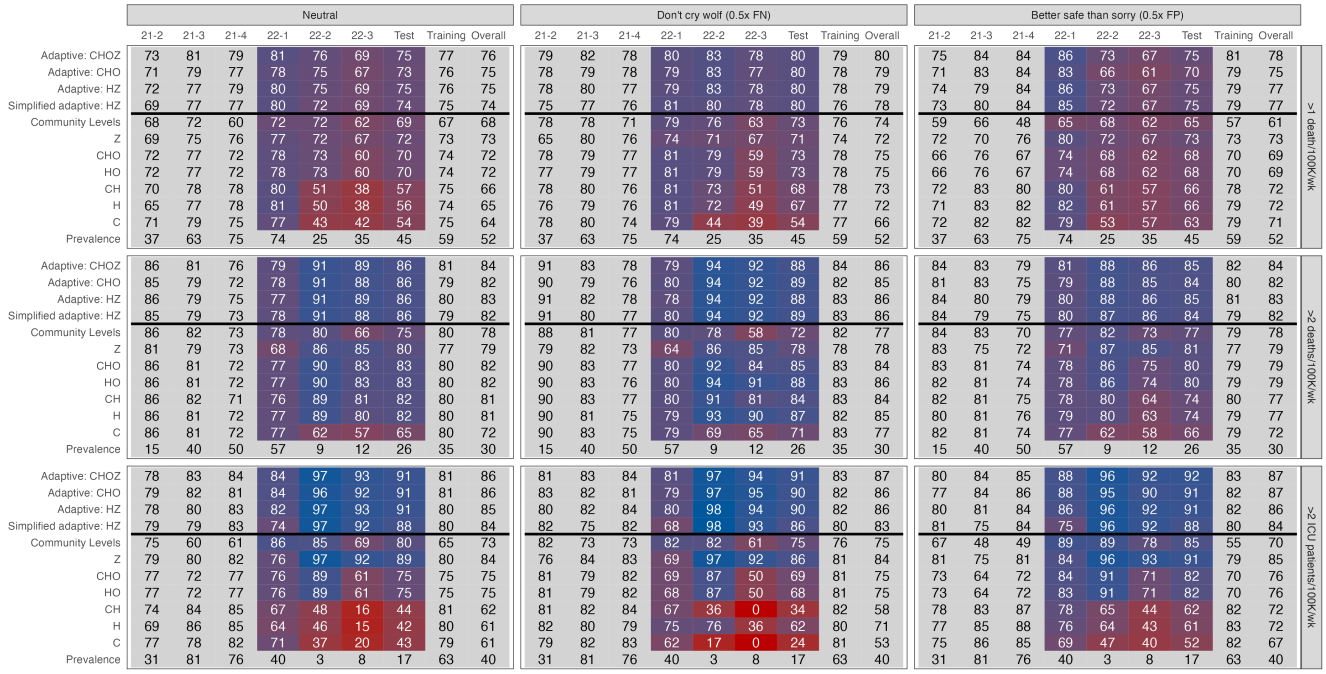


Fig. S7. County-level results by quarter. Metrics are displayed on the left, with training data from Q3-Q4 2021 and test data from Q1-Q3 2022. Cells report weighted accuracy. Preferences for false positive versus false negatives varying across columns (with "neutral" corresponding to unweighted accuracy) and outcomes across rows. Prevalence indicates the proportion of high location weeks in a given quarter.

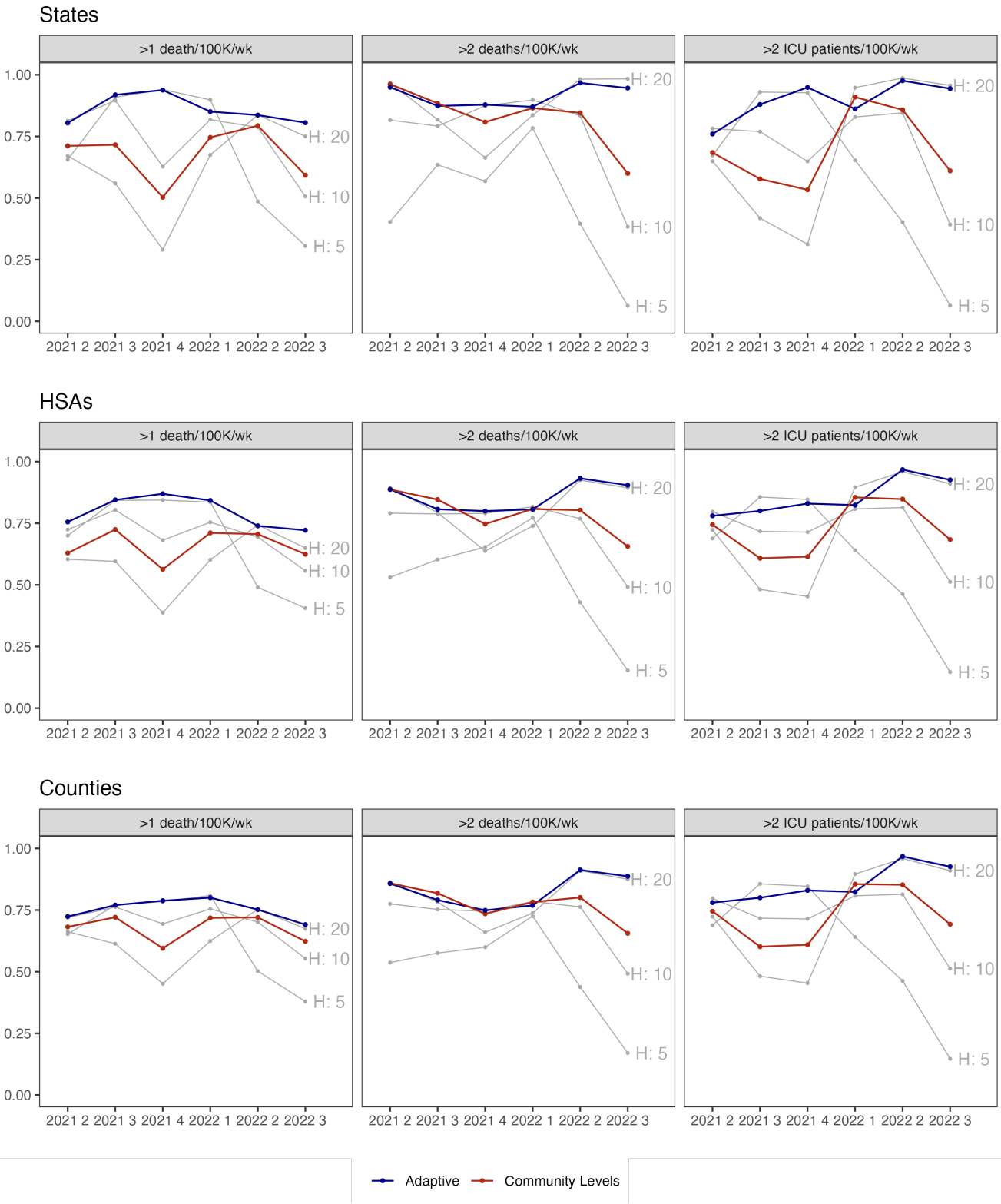


Fig. S8. Weighted accuracy by metric. Columns indicate different outcomes. The x-axis indicates quarter, and the y-axis predictive accuracy. Grey lines depict metrics based on new hospital admissions exceeding the row threshold. The red line indicates CDC Community Level and the blue line an adaptive metric (HZ).

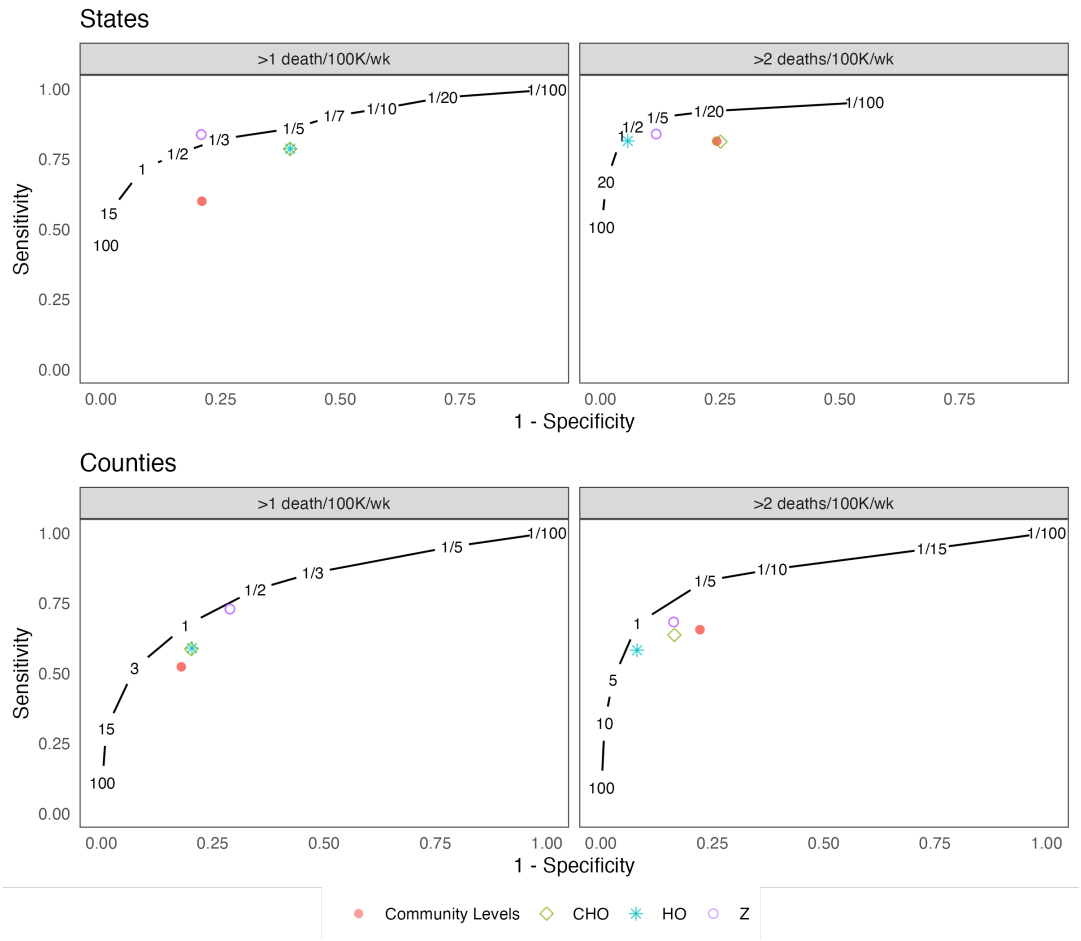
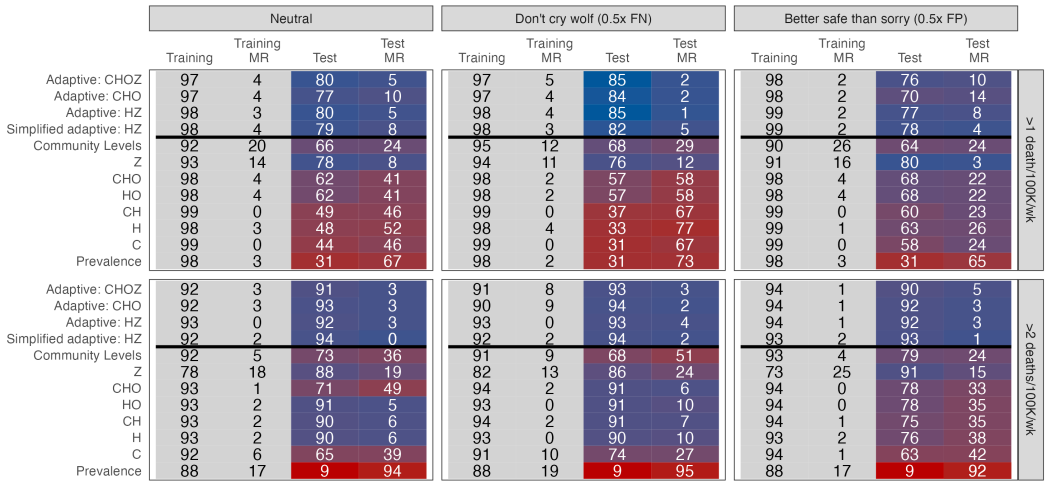
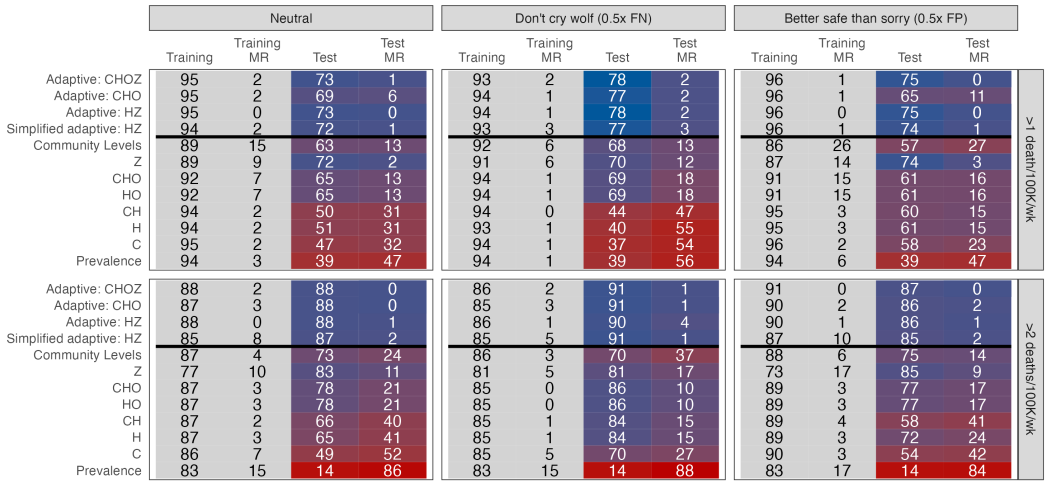


Fig. S9. Receiver operating characteristic (ROC) curves for the test period from January 1, 2022 to September 30, 2022. The black line indicates performance of the adaptive metric (HZ) across different values of wt, indicating the relative preference for false negatives over false positives. The top plot indicates states and the bottom plot counties.

States



HSAs



Counties

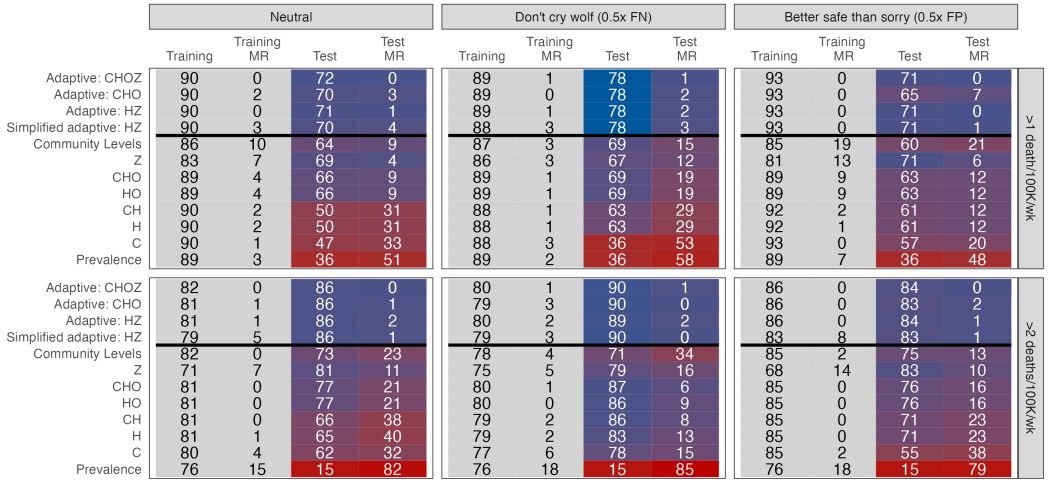


Fig. S10. Head-to-head comparison results (omicron training set). The top plots display results from state-level analyses and the bottom plots display results from county-level analyses, both weighted for population. Metrics are displayed on the left, with training data from December 15, 2021-February 15, 2022 and test data from February 16-September 30, 2022. Cells report weighted accuracy and maximum regret (MR) over training and test periods. Rows vary outcomes, and columns vary preferences for false positive versus false negatives, with "neutral" corresponding to unweighted accuracy. Prevalence indicates the proportion of high location-weeks in a given time period.

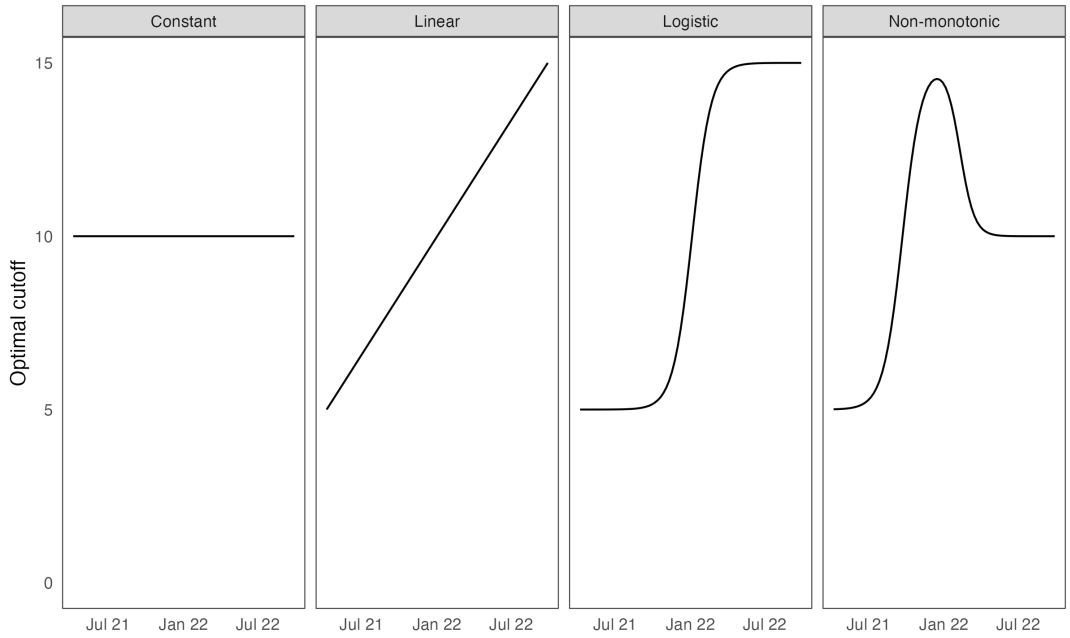


Fig. S11. Simulation scenarios. We vary the optimal cutoff for hospitalization to classify a location-week as "high" ($-\beta_0/\beta_1$) over time in different scenarios. The constant scenario assumes a static relationship between indicators and outcomes, while other scenarios assume a changing relationship.

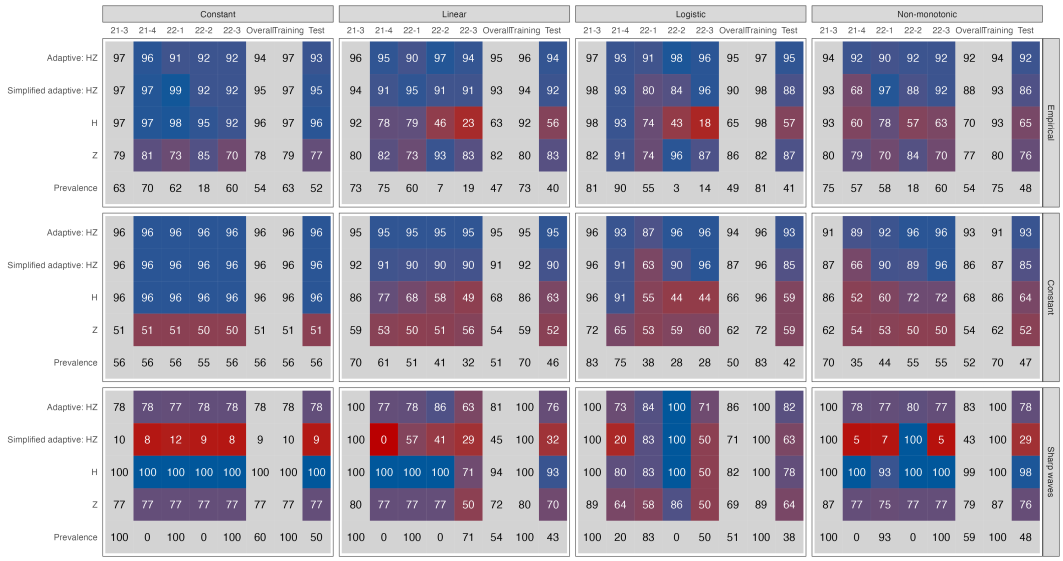


Fig. S12. Simulation results. Columns vary the relationship between the input indicator and outcome over time (Figure S11) and rows vary indicator (input) prevalence. Metrics are varied over the y-axis, and 3-week-ahead predictive accuracy is displayed in cells.