PNAS RESEARCH ARTICLE | MEDICAL SCIENCES SOCIAL SCIENCES OPEN ACCESS

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

Alyssa M. Bilinski[a,1], Joshua A. Salomon[b], and Laura A. Hatfield[c]

Edited by Larry Wasserman, Carnegie Mellon University, Pittsburgh, PA; received February 13, 2023; accepted April 27, 2023

Throughout the COVID-19 pandemic, policymakers have proposed risk metrics, such as the CDC Community Levels, to guide local and state decision-making. However, risk metrics have not reliably predicted key outcomes and have often lacked transparency in terms of prioritization of false-positive versus false-negative signals. They have also struggled to maintain relevance over time due to slow and infrequent updates addressing new variants and shifts in vaccine- and infection-induced immunity. We make two contributions to address these weaknesses. We first present a framework to evaluate predictive accuracy based on policy targets related to severe disease and mortality, allowing for explicit preferences toward false-negative versus false-positive signals. This approach allows policymakers to optimize metrics for specific preferences and interventions. Second, we propose a method to update risk thresholds in real time. We show that this adaptive approach to designating areas as "high risk" improves performance over static metrics in predicting 3-wk-ahead mortality and intensive care usage at both state and county levels. We also demonstrate that with our approach, using only new hospital admissions to predict 3-wk-ahead mortality and intensive care usage has performed consistently as well as metrics that also include cases and inpatient bed usage. Our results highlight that a key challenge for COVID-19 risk prediction is the changing relationship between indicators and outcomes of policy interest. Adaptive metrics therefore have a unique advantage in a rapidly evolving pandemic context.

infectious disease dynamics | decision theory | risk prediction | COVID-19

Significance

In the rapidly evolving COVID-19 pandemic, public health risk metrics have often become less relevant over time. Risk metrics are designed to predict future severe disease and mortality based on currently available surveillance data, such as cases and hospitalizations. However, the relationship between cases, hospitalizations, and mortality has varied considerably over the course of the pandemic, in the context of new variants and shifts in vaccine- and infection-induced immunity. We propose an adaptive method for risk designations that is regularly updated to reflect the evolving relationship between surveillance data inputs and future outcomes of policy interest. Our method captures changing pandemic dynamics, requires only hospitalization input data, and outperforms static methods, providing more reliable and actionable risk designations.

Understanding the evolution of infectious disease risk is critical for individuals making decisions about personal precautions, policymakers recommending mitigation measures, and health care institutions planning for future surges. Throughout the COVID-19 pandemic, indicators such as reported cases and percent of PCR tests positive for SARS-CoV-2 have been used to guide pandemic response (1–4). Currently, the Centers for Disease Control and Prevention (CDC)'s Community Levels designate areas as low, medium, or high risk based on reported cases, new COVID-19 hospital admissions, and the percentage of inpatient beds occupied by COVID-19 patients (2).

However, COVID-19 risk metrics have several weaknesses. First, policymakers have struggled to identify leading indicators of key health outcomes. For example, PCR test positivity was abandoned as a trigger for school closures because it did not reliably predict in-school transmission (5). Similarly, Community Transmission metrics developed by the CDC based on cases and test positivity were deemphasized due to poor prediction of future severe outcomes (2). Other community metrics have focused on predicting severe disease and mortality (2, 6). For example, the indicators used in CDC Community Levels were selected because they correlated with ICU rates and mortality 3 wk in the future (2). However, the thresholds for low, medium, and high were not selected to correspond to specific future mortality rates (7), thus complicating the understanding of a high-risk designation.

Second, many metrics fail to distinguish different error types. Falsely classifying an area as high risk may prompt unnecessary or harmful interventions, while a false negative may fail to activate needed public health measures (8). Individuals and policymakers may vary in their preferences for avoiding these two types of errors, but current methods fail even to make these preferences explicit (9).

Finally, changes in available data, COVID-19 variants, and levels of immunity can render metrics obsolete as the pandemic evolves (10). For instance, with the omicron variant, cases and hospital admissions have corresponded to lower levels of mortality than in earlier waves. Shifts from PCR to at-home testing and changes in case reporting have also made case data less reliable and available over time (11, 12). Ad hoc updates to risk designations are insufficient to ensure that the metrics remain relevant. Moreover,

Author contributions: A.M.B. and J.A.S. designed research; A.M.B. performed research; A.M.B. and L.A.H. analyzed data; and A.M.B., J.A.S., and L.A.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: alyssa_bilinski@brown.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2302528120/-/DCSupplemental.

Published August 1, 2023.

transparency in the process is key to alleviating concerns about "moving the goalposts" (13).

This paper makes two contributions to address these weaknesses in the context of COVID-19 community risk metrics. First, we propose a framework for predictive accuracy that incorporates preferences over false negatives versus false positives, using weights to optimize metrics for specific policy objectives. Second, we present a method to update risk thresholds over time and show that this adaptive approach outperforms static metrics. With our approach, we demonstrate that metrics using only new hospital admissions often perform as well in prediction as metrics that also include cases and inpatient bed usage.

Materials and Methods

The CDC used indicators available nationwide (cases, hospitalizations, and occupancy of staffed inpatient beds) to develop Community Levels (2). In this research, we used the same indicators to define alternative state and county metrics, then compared metrics based on their ability to predict future health outcomes.

Outcomes. The primary evaluation criterion was predictive power for high mortality. We defined "high mortality" as >1 death per 100,000 per week and "very high mortality" as >2 deaths per 100,000 per week. The lower threshold was defined in reference to peak mortality of other respiratory viruses (influenza and respiratory syncytial virus) during a severe season (7, 14). Let $T \in 1, 2$ denote these mortality thresholds. The true outcome was a binary variable equal to 1 if mortality three weeks from the current week (i.e., at time $w+3$) in location i exceeded the threshold; formally, $Y_{i,w+3} = \mathbb{I}(\text{mortality at } w+3 > T) \in 0, 1$. In secondary analyses of health care strain, we evaluated predictive power for 3-wk-ahead ICU admissions, defining "high" as >2 ICU hospitalizations per 100,000 population per week, and as >10% for 3-wk-ahead COVID-19 inpatient bed occupancy (the lowest threshold meeting the CDC classification of "high inpatient bed usage") (2).

We used a 3-wk prediction window because previous CDC analyses indicated that this maximized the correlation between indicators and severe outcomes (2). This also reflects the necessary lead-time for interventions to begin to have an impact on severe outcomes; a metric that predicts severe mortality tomorrow will come too late for effective action. We used discrete outcomes to mirror CDC risk categories and to reflect the common practice of adopting pandemic interventions in response to threshold crossing.

Indicators. Indicators are the observed quantities that enter our prediction models. We used the same three indicators as the CDC's Community Levels: new COVID-19 cases per 100,000 (weekly total), new COVID-19 hospital admissions per 100,000 (weekly total), and the occupancy of staffed inpatient hospital beds by COVID-19 patients (7-d average). Let $X_{C,i,w}$, $X_{H,i,w}$, and $X_{O,i,w}$ denote the levels of these three indicators respectively, in location i during week w.

Data. We obtained data on indicators and outcomes at both state and county levels and conducted separate analyses for each geographic level. For cases and deaths, we used aggregated counts compiled by state and local health agencies (15). For new COVID-19 admissions and bed occupancy, we used data reported to the US Department of Health and Human Services Unified Hospital Data Surveillance System (16, 17). Consistent with CDC Community Level calculations, we calculated county-level hospitalizations at the Health Service Area (HSA)-level to account for care-seeking across counties and computed measures at the midpoint of each week (2). HSAs are defined by the National Center for Health Statistics to be one or more contiguous counties with self-contained hospital care (18). In sensitivity analyses, we also present analyses with all inputs and outcomes calculated at the HSA-level.

Metrics. Metrics take indicators as inputs and produce a binary risk classification for a geographic area as output. Our metrics used data available at week w to predict outcomes above the prespecified threshold, T, 3 wk in the future, classifying a locality as high risk, $\hat{Y}_{w+3} = 1$, or not high-risk, $\hat{Y}_{w+3} = 0$. (For readability, we omit location subscripts i when referring to a single observation in this section.)

Objective. We used weighted classification accuracy to compare metrics on their ability to predict future outcomes, where weights reflected preferences for avoiding different types of errors.

We assumed a simple underlying decision-analytic framework: a decision maker receives a prediction of, for example, mortality 3 wk hence, $\hat{Y}_{w+3}$, and takes action in response to that prediction. If the metric predicts high mortality ($\hat{Y}_{w+3} = 1$), she will take one action; if the model does not predict high mortality ($\hat{Y}_{w+3} = 0$), she will take a different action. Each action has benefits and costs that depend on the true outcome. For example, avoiding unnecessary interventions under a true negative conserves public health resources, while inaction due to a false negative may lead hospitals to become overburdened. By contrast, a false positive may have costs such as wasted resources and harming public trust due to unnecessary interventions.

We consider costs in terms of disease burden and public health resources. We anchor costs at 0 in the scenario in which the model correctly predicts low mortality ($\hat{Y}_{w+3} = Y_{w+3} = 0$). If the model incorrectly predicts high mortality ($\hat{Y}_{w+3} = 1$, $Y_{w+3} = 0$), we denote public health resources spent and social costs as S_0. By contrast, if a model incorrectly predicts low mortality ($\hat{Y}_{w+3} = 0$, $Y_{w+3} = 1$), policymakers incur disease costs of D. Last, if a model correctly predicts high mortality ($\hat{Y}_{w+3} = Y_{w+3} = 1$), we assume policymakers implement an intervention that reduces disease by a factor of α, but pay resource costs, for a total cost of $(1-\alpha)D + S_1$.

The total cost associated with a particular metric, M (omitting subscripts for parsimony) is:

$$\begin{aligned}C(M) &= Pr(\hat{Y}=1, Y=0)S_0 + Pr(\hat{Y}=0, Y=1)D \\ &\quad + Pr(\hat{Y}=1, Y=1)\left((1-\alpha)D + S_1\right) \\ &= Pr(\hat{Y}=1, Y=0)S_0 \\ &\quad + Pr(\hat{Y}=0, Y=1)(\alpha D - S_1) \\ &\quad + Pr(Y=1)((1-\alpha)D + S_1).\end{aligned}$$

Because the last term is constant across all metrics (which cannot affect prevalence of high outcomes), this cost is proportional to the weighted misclassification rate:

$$\begin{aligned}C(M) &\propto p_{FP}S_0 + p_{FN}(\alpha D - S_1) \\ &\propto p_{FP} + p_{FN}wt.\end{aligned}$$

We can therefore rank metrics based only on performance (i.e., their probabilities of making each error type) and the decision maker's relative preference for false positives compared to false negatives (wt). As the above expression indicates, we can conceptualize weight wt as the ratio of the net benefit from taking action on a true positive ($\alpha D - S_1$) to costs incurred by unnecessary action in the case of a false positive (S_0).

We considered three values of this weight. "Neutral" weighted false negatives and false positives equally ($wt = 1$, equivalent to unweighted accuracy), "don't cry wolf" down-weighted false negatives as half the cost of false positives ($wt = 0.5$), and "better safe than sorry" down-weighted false positives as half the cost of false negatives ($wt = 2$).

We estimated weighted accuracy for each metric as 1 minus the weighted misclassification rate:

$$\delta_{wt}(M) = 1 - p_{FP}w_P - p_{FN}w_N.$$

While any w_N and w_P such that $\frac{w_P}{w_N} = wt$ would produce the same ranking of metrics, the absolute value of δ_{wt} depends on w_N and w_P. We set w_N and w_P such that both error weights are shifted equally in magnitude to achieve the desired ratio, with an increase in one and corresponding decrease in the other. That is, we set w_N and w_P using the value a such that $w_N = (1 - a)$, $w_P = (1 + a)$, and $w_N/w_P = (1 - a)/(1 + a) = wt$. With neutral weighting, $w_N = w_P = 1$.

We used weighted accuracy as our primary measure of performance, with higher weighted accuracy indicating better performance. We further weighted δ_{wt} by population to reflect the total proportion of individuals living in a location with an accurate classification (*SI Appendix*, Text A).

Static metrics. We considered two types of metrics, static and adaptive. Static metrics used the same threshold each week to classify a locality as high risk. They differed in input indicators, which could include 1) new cases only (C), 2) new hospital admissions only (H), 3) cases and hospital admissions (CH), 4) hospital admissions and bed occupancy (HO) or 5) all three indicators (CHO)). We varied the threshold on cases from 50 to 300 per 100,000 (in increments of 50), on new hospitalizations from 5 to 25 per 100,000 (in increments of 5), and on occupancy from 5 to 20% (in increments of 5). We designated an area as high risk if all the indicators in a given indicator set were above their specified thresholds.

We also replicated the CDC's Community Levels, designating an area as highrisk if

$$[X_{C,i,w} < 200 \text{ AND } (X_{H,i,w} \geq 20 \text{ OR } X_{O,i,w} \geq 15\%)] \text{ OR}$$
$$[X_{C,i,w} \geq 200 \text{ AND } (X_{H,i,w} \geq 10 \text{ OR } X_{O,i,w} \geq 10\%)].$$

Last, we considered a metric (Z) that designated an area as high risk if the outcome was above the threshold of interest at the time of prediction, i.e., $\hat{Y}_{i,w+3} = \mathbb{I}(Y_{i,w} = 1)$, predicting $\hat{Y}_{i,w+3} = 1$ at time $w + 3$ only if $Y_{i,w}$ was equal to 1, indicating the area was currently observing the high designation.

Adaptive metrics. Adaptive metrics changed thresholds over time based on their ability to predict mortality during the recent past (Fig. 1). At time w, we used the most recent weeks of past indicator data with complete 3-wk-ahead outcomes as training data. To these training data, we fit logistic regression models with outcomes on the *Left*-hand side and indicators from previous weeks on the *Right*-hand side. For example, in the model corresponding to the CHO indicator set, we fit

$$\text{logit}(Pr(Y_{i,v} = 1)) = \beta_0 + \beta_1 X_{C,i,v-3} + \beta_2 X_{H,i,v-3} + \beta_3 X_{O,i,v-3}. \quad \textbf{[1]}$$

for $v \in [w - 3, w]$. From this model, we obtained $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, which we then used to produce fitted probabilities for each locality's mortality 3 wk ahead using:

$$\widehat{Pr}(Y_{i,w+3} = 1) = \text{logit}^{-1}\left(\hat{\beta}_0 + \hat{\beta}_1 X_{C,i,w} + \hat{\beta}_2 X_{H,i,w} + \hat{\beta}_3 X_{O,i,w}\right). \quad \textbf{[2]}$$

Logistic regression smoothed over noise in the small training data and reduced the dimension of multiple indicators by converting to a probability scale.

With predictions on a probability scale, we specified a probability cutoff above which we classified a location as high risk. We selected this cutoff based on the relative weighting of different error types (wt). We classified a locality as high risk whenever the probability was above $1/(1 + wt)$ (*SI Appendix*, Text B for optimal cutoff derivation). For our three weights (neutral, don't cry wolf, and better safe than sorry), the cutoff values were $\frac{1}{2}$, $\frac{2}{3}$, and $\frac{1}{3}$, respectively. With a single predictor, this process would be equivalent to identifying the optimal threshold for the indicator over the training period, accounting for user preferences.

To assess sensitivity to different functional forms, we specified analogous models based on CHOZ and HZ indicator sets and an additional model (CHOD) that included all indicators as well as the change in each indicator from the prior week. We also included a simplified version that was updated less frequently, only refitting to the training data each quarter, rather than each week. We varied the number of training weeks from 4 to 12 (i.e., fitting Eq. **1** to training datasets as large as $v \in [w - 11, w]$).

Head-to-Head Comparison. We compared the performance of the metrics during training and out-of-sample test periods. To

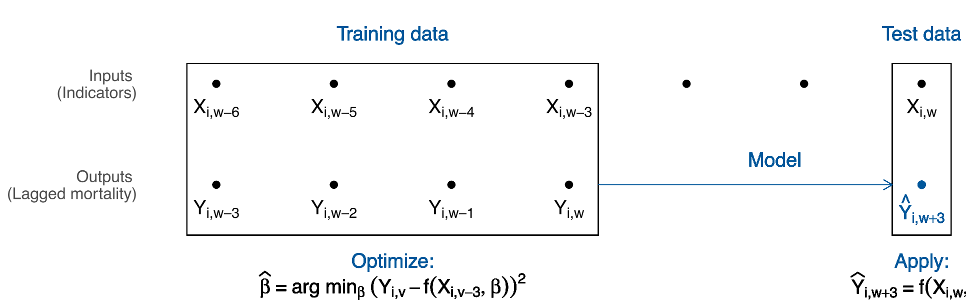

Fig. 1. Adaptive metrics. The diagram shows the model-fitting process using 4 wk of training data. We trained a model using the 4 most recent weeks with complete outcome data, including inputs from $w - 6$ to $w - 3$ and outputs from $w - 3$ to w. We then used this model, with input data from w, to estimate the probability of "high" or "very high" future mortality at $w + 3$ and specified a binary prediction based on whether this probability exceeded the user's cutoff. (When a single indicator is used as the only input, this process is equivalent to identifying the optimal threshold for the indicator over the training period, accounting for user preferences.)

define the training period, we began with the window the CDC used to fit Community Levels (March 1, 2021, through January 24, 2022). We further allowed the month of March for model fitting including 3 wk of past mortality data. Thus, our training inputs spanned April 1, 2021 through December 31, 2021, 2021 Q3 and Q4, with outcomes extending through January 21, 2022.

We compared performance across metrics separately for each outcome (e.g., >1 or >2 deaths/100k/wk), preference weight ($wt = 0.5$, 1, or 2), and geographic area (state or county). Within each combination of these, we chose the best-performing static metric during the training period from among the 6, 5, 30, 20, or 120 possibilities within the C, H, CH, HO, and CHO indicator sets. The CDC Community Levels and current outcome (Z) metrics were fixed, so there was no selection within this metric type. For adaptive metrics, we used the training period to optimize the number of training weeks.

Performance evaluation. We present weighted accuracy of each selected metric in the training quarters (during which the best performer of each type was selected) and a test period of January 1, 2022, through September 30, 2022 (i.e., 2022 Q1–Q3). As a sensitivity analysis, we used December 15, 2021 through February 15, 2022, as a training period, to only include training data from the omicron period when the infection-fatality rate fell sharply. We then used data from February 16 through September 30, 2022, as the test period.

In addition to presenting overall weighted accuracy, we summarize variation in performance across quarters with maximum quarterly regret, the difference between a metric's predictive accuracy and the best performing metric (19). We calculated regret for each selected metric in each quarter and took the maximum across quarters:

$$MR_M = \max_{q \in \mathcal{Q}} \left(\max_{m \in \mathcal{M}} \delta_{wt,q}(m) \right) - \delta_{wt,q}(M),$$

where M is a metric of interest, $\mathcal{Q}$ is a set of quarters, $\mathcal{M}$ is a set of metrics, and $\delta_{wt,q}$ is weighted accuracy during quarter q.

Last, to decompose variation between metrics into differences in predictive power and differences in error preferences, we computed sensitivity ($Pr(\hat{Y}_{i,w+3} = 1 | Y_{i,w+3} = 1)$) and specificity ($Pr(\hat{Y}_{i,w+3} = 0 | Y_{i,w+3} = 0)$) across different wt values for adaptive metrics and compared these to sensitivity and specificity for static metrics.

Simulations. To generalize our approach beyond the specific pandemic periods considered, we developed simple simulations, varying the relationship between indicators and outcomes over time as well as the prevalence of high maturity outcomes (*SI Appendix*, Text C). We considered several functional forms for the relationship between inputs and synthetic outputs, including a scenario with a true constant optimal cutoff above which to classify $\hat{Y}_{w+3}$ as 1 and scenarios with time-varying optimal cutoffs (linear, logistic, and nonmonotonic). We also varied prevalence of high mortality outcomes, including a constant case, a case based on empirical hospitalization waves, and a case in which waves designed to be much sharper than true waves. We then estimated predictive accuracy across different scenarios.

Results

Indicator levels and mortality varied substantially over the study period (Fig. 2), which included two major waves of illness (delta and omicron BA.1) and a smaller wave in summer 2022 (omicron BA.5) (*SI Appendix*, Figs. S2 and S3 for detailed dynamics of indicators and outcomes over the study period.) The percentage of population-weighted state-weeks with high future mortality ranged from a peak of 94% during Q4 2021 to a low of 17% during Q2 2021. For very high mortality, this ranged from 61% (Q1 2022) to 3% (Q2 2022). We observed similar variation in counties, with less extreme swings (e.g., from 74% to 25% for high mortality). The relationship between indicators and outcomes shifted substantially over the period studied. In particular, in the third quarter of 2022, cases, hospitalizations, and bed occupancy all increased, but mortality remained lower than in previous waves (Fig. 2).

Static Metrics. In Fig. 3, we present the performance of the best-performing static metrics from different health care indicator sets (C, H, CH, HO, and CHO) during the training and test periods. Recall that the static metrics designated an area as high-risk if all included indicators exceeded their respective optimal thresholds from the training period. For the high mortality outcome (>1 death/100k/wk) at the state level with neutral weighting, the chosen thresholds for static metrics were 50 cases/100k (C); 5 hospitalizations/100k (H); 50 cases/100k, 5 hospitalizations/100k (CH); 5 hospitalizations/100k, 5% bed occupancy (HO); and 50 cases/100k, 5 hospitalizations/100k, 5% bed occupancy (CHO). The thresholds for the remaining outcomes and geographic levels are given in *SI Appendix*, Table S1.

During the training period, there were only minor differences in training accuracy between static metrics that used different health care indicator sets (e.g., 83 to 87% in predicting high mortality for states with neutral weighting, 73 to 75% for counties). However, for nearly all static metrics and outcomes, test accuracy was lower and more variable than training accuracy (e.g., 45 to 68% and 54 to 70% for high mortality in states and counties, respectively).

Some of this variation was due to the shifting relationship between indicators and lagged outcomes over time. We illustrate this in Fig. 4, where gray lines show the performance of metrics based on different hospitalization cutoffs with neutral weighting. No single cutoff dominated during the full study period. For example, the cutoff of 5 per 100,000 performed best for high mortality during 3 quarters of the study period, with accuracy above 90% in states and 75% in counties, but was the worst performing in Q2–Q3 2022, with less than 50% accuracy. The accuracy of the single best-performing metric also varied across quarters (e.g., from 61 to 80% for high mortality and 72 to 90% for very high mortality in counties).

Other static metrics similarly reflected the evolving relationship between indicators and mortality. For example, while prediction based on current risk designation (Z) was the second-worst performing static indicator during the training period for high mortality (after Community Levels) in states, it performed best during the test period, when waves of infection were less extreme and variable. CDC Community Levels performed relatively worse compared to other static metrics at predicting high mortality during the training period, but similar or better during the test period; the converse was true for predicting very high mortality (Fig. 3). Overall, static metrics that used hospitalizations and bed occupancy (HO) performed most consistently across training and test periods, but we would have been unable to discern this with only training data. Across static metrics, training accuracy was an unreliable signal of test accuracy.

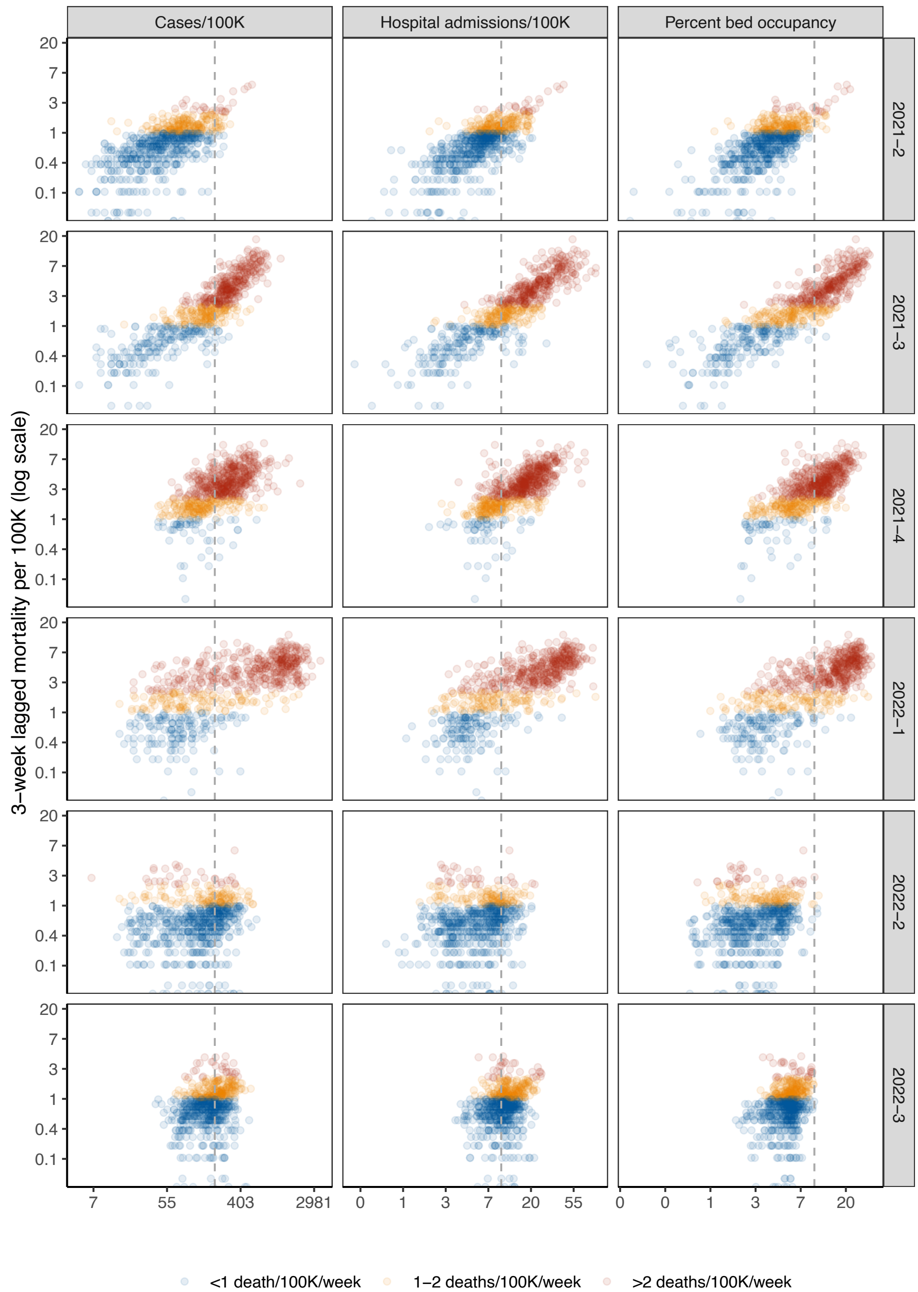

Fig. 2. State-level lagged mortality vs. indicator levels by quarter. Columns indicate different indicators (weekly cases per 100,000 population, new hospital admissions per 100,000 population, and percentage of inpatient beds occupied by COVID-19 patients), and rows indicate quarters. The x-axis displays indicator values on a log scale, and the y-axis displays 3-week-ahead mortality per 100,000 population on a log scale. Each point on the scatterplot is a state-week. Colors show the mortality outcome level. The vertical gray dotted lines indicate thresholds from CDC Community Levels for each indicator (≥200 cases/100K/wk and ≥10 new admissions/100K/wk or ≥10% COVID-19 bed occupancy). See *SI Appendix*, Fig. S1 for a county-level plot.

Adaptive Metrics. Adaptive metrics consistently outperformed static metrics for both primary outcomes in training and test periods (Fig. 3). For example, when predicting high mortality in states with neutral weighting, adaptive metrics had an overall accuracy of 86 to 89% in the training period and 77 to 83% in the test period; for very high mortality, this was 85 to 90% and 91 to 94% respectively. While all adaptive functional forms performed well, metrics corresponding to CHOZ and HZ (88 to 89% training, 83% test for high mortality) slightly outperformed CHO and the simplified HZ version with less frequent updating. They also performed better than metrics that included week-on-week indicator changes (CHOD). Importantly, while adaptive metrics performed similarly to static metrics during some quarters, they rarely underperformed by a substantial margin and often achieved substantial gains (Fig. 4). This was reflected in regret, which was better controlled by adaptive metrics than static metrics in nearly all cases at both state and county levels. Adaptive metrics also weakly dominated static indicator-based metrics and Community Levels in the sense that HZ could achieve at least equal (and often higher) sensitivity and specificity for at least one value of *wt* at both geographic levels (*SI Appendix*, Fig. S11).

Alternative Preferences, Secondary Outcomes, and Sensitivity Analyses. Adaptive metrics similarly outperformed static metrics

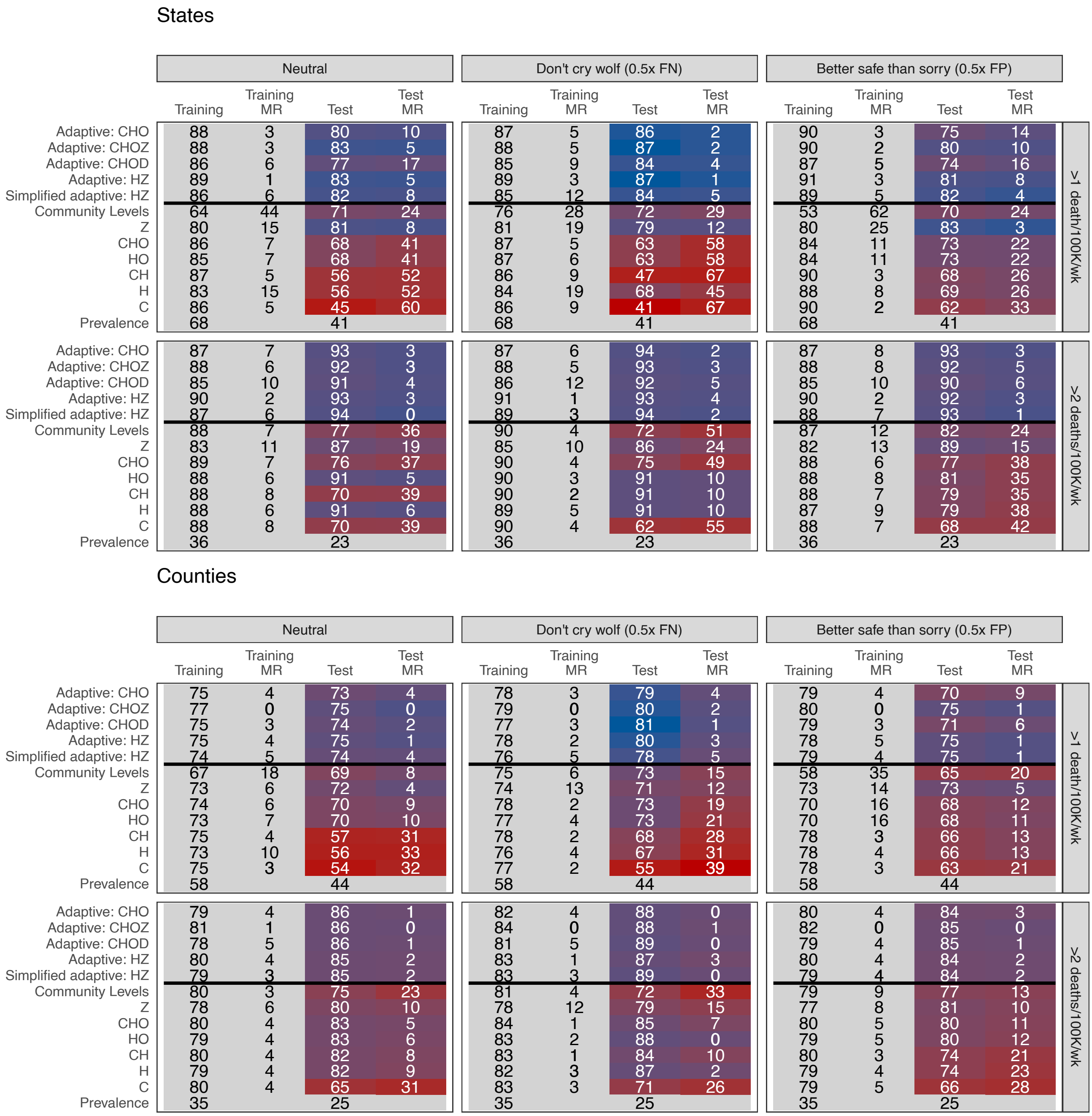

Fig. 3. Head-to-head comparison results. The top plots display results from state-level analyses and the bottom plots display results from county-level analyses, both weighted for population. Metrics are displayed on the *Left*, with training data from Q2–Q4 2021 and test data from Q1–Q3 2022. Cells report weighted accuracy and maximum regret (MR) over training and test periods. Rows vary outcomes, and columns vary preferences for false positives versus false negatives, with "neutral" corresponding to unweighted accuracy. Prevalence indicates the proportion of high location-weeks in a given time period. A version including HSA-level analyses can be found in *SI Appendix*, Fig. S4. Secondary outcomes are presented in *SI Appendix*, Fig. S5, and weighted accuracy by quarter is presented in *SI Appendix*, Figs. S6–S8. For adaptive metrics, models vary functional form to include: 1) CHO (cases, hospitalizations, inpatient bed occupancy); 2) CHOZ (cases, hospitalizations, inpatient bed occupancy, current risk designation); 3) CHOD (cases, hospitalizations, inpatient bed occupancy, weekly changes in each indicator); 4) HZ (hospitalizations, current risk designation); 5) Simplified HZ (hospitalizations, current risk designation—updated quarterly). (For additional adaptive functional forms, *SI Appendix*, Fig. S9.)

across preference weights (Fig. 3) and for secondary outcomes of future ICU hospitalizations over 2 per 100,000 and future COVID-19 inpatient bed occupancy >10% (Fig. 4 and *SI Appendix*, Fig. S5). Across outcomes, we only observed substantial improvement in predictive performance from adding weekly changes for the inpatient bed occupancy outcome; for this outcome, adaptive metrics without weekly changes had smaller improvements over static metrics (*SI Appendix*, Figs. S6–S8).

The gain in weighted accuracy for adaptive metrics was higher when estimated at the HSA level rather than at the county level (about 2 percentage points for both mortality outcomes with neutral weighting) (*SI Appendix*, Fig. S4). Running the training period from December 15 to February 15 to capture the omicron variant did not substantially alter the relative benefit of adaptive metrics, with a 14 percentage point increase in weighted accuracy in states for high mortality compared to Community Levels with

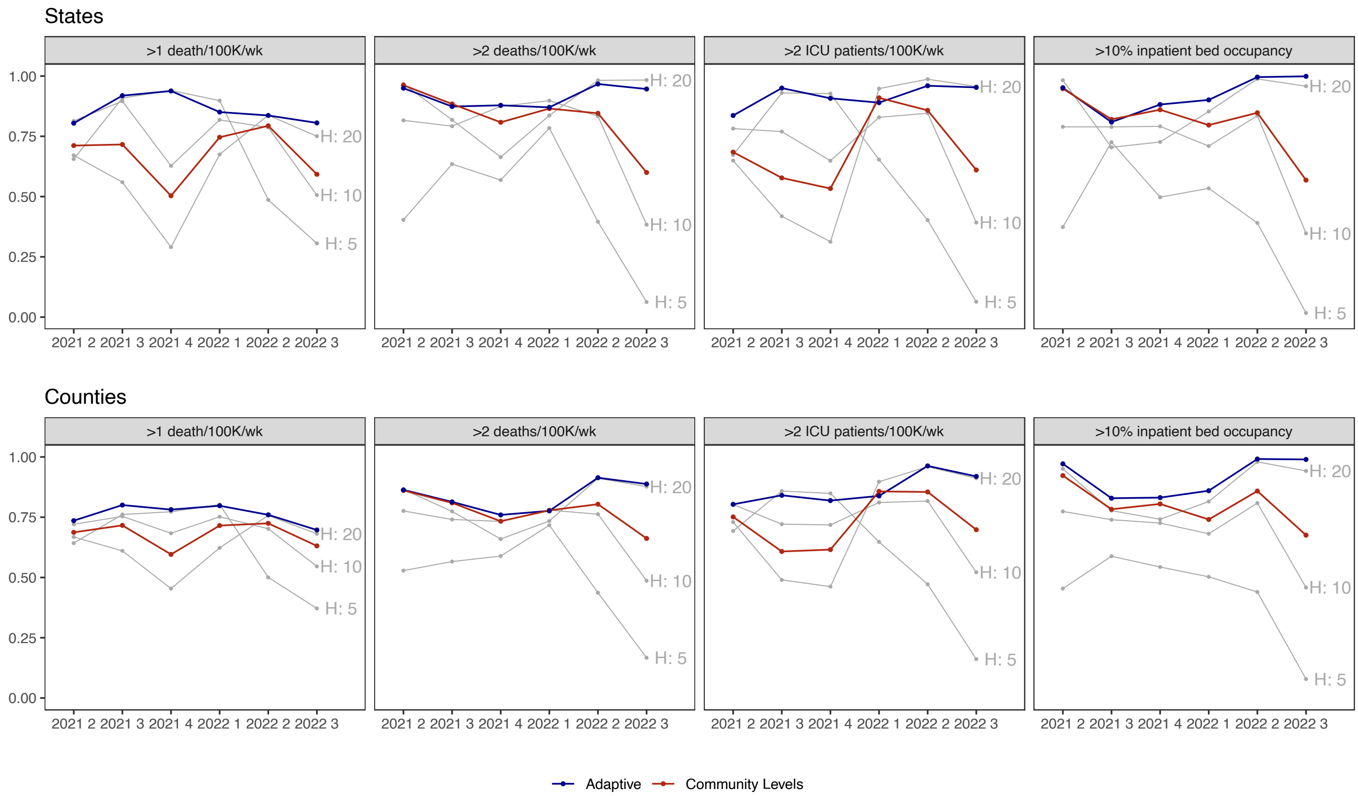

Fig. 4. Weighted accuracy by metric. The top plot displays states, and the bottom plot displays counties. Columns indicate different outcomes. The x-axis indicates quarter, and the y-axis predictive accuracy with neutral weighting. Gray lines depict metrics based on new hospital admissions exceeding the labeled threshold. The red line indicates CDC Community Levels and the blue line the best-performing adaptive metric in the training period of those listed in Fig. 3. A version with HSA-level results can be found in *SI Appendix*, Fig. S10.

a neutral weighting (compared to 12% in the base case) and 7% in counties (compared to 6%) (*SI Appendix*, Fig. S12).

Simulations. In simulations, adaptive metrics outperformed static metrics when the relationship between indicators and outcomes was changing over time, across different input/output functional forms and regardless of whether prevalence was constant or followed waves generated from empirical hospitalization data (*SI Appendix*, Fig. S14). There was no gain when the relationship between indicators and outcomes was constant; adaptive metrics performed worse than static metrics when waves were extremely sharp, and there could be insufficient training data near the threshold to estimate the optimal cutoff.

Discussion

We proposed an adaptive approach to estimating local risk which continually updates metrics to ensure they predict outcomes of policy interest. We showed that this would have outperformed static approaches, including CDC Community Levels over the past year. Our metrics have a unique advantage in a rapidly evolving pandemic context. They quickly pick up new information as the relationship between indicators and future mortality shifts, allowing us to refine the threshold for "high risk" and improve discrimination.

Previous papers have proposed adaptive policies for COVID-19 management, in which policymakers shift responses depending on observed indicators like cases and deaths (20–22). We extend this work by allowing the trigger thresholds for indicators to also vary over time. Such an approach could be particularly advantageous for maintaining public trust when the relationship between indicators and outcomes is not yet well-understood or is changing quickly (23).

Our approach draws on ideas that have been applied in the online calibration literature and in forecasting, but have not yet been widely applied to population risk metrics (6, 24–26). Nevertheless, some previous authors have noted that accounting for the evolving pandemic conditions is important for effective decision-making, suggesting policies that are adjusted for the changing costs of mitigation over the course of a pandemic or the number of people vaccinated over time (27, 28). We particularly emphasize parsimony for policy metrics, demonstrating that policymakers can obtain equal predictive performance with fewer inputs potentially reducing the burden of data collection on state and local public health departments. Similar to other authors, we find hospitalizations to be the most powerful predictor of future mortality (6). We further emphasize that it is valuable to collect real-time data on outcomes of policy interest, like mortality. In the case of COVID-19, while state mortality is still collected and reported weekly, many counties have reduced reporting frequency (15).

Our method can also reflect a policymaker's preferences for the trade-off between avoiding false negatives and false positives, filling a previously identified gap between models and decision theory (29). In practice, different indicators could be used to guide different policies. For the most burdensome interventions (e.g., business closures), policymakers might prefer a low risk of false negatives, while for less burdensome interventions, (e.g., distribution of rapid tests), they might have a higher tolerance for false positives. Future work could formally expand adaptive metrics to include multiple levels of risk designations (e.g., low/medium/high) based on different outcomes of interest, prediction of multiple levels of a single outcome, or different preferences for false negatives versus false positives. Metrics could also be modified to reflect different outcomes for different users, such as employers and workplaces, and to map designations to institution-specific risk tolerances.

There are several additional limitations and potential extensions to this study. First, we model only outcomes related to severe disease and death from COVID-19, as national policymakers have designated these priority outcomes. Nevertheless, metrics to track illness are also important for understanding the full burden of disease, which can include disruptions from illness and Long COVID, and work is also needed to predict surges with longer lead time (26, 30). In addition, no adaptive framework can automatically incorporate all possible variations. Manual tuning may be needed, for example, if the frequency of reporting of hospitalization changes over time. Furthermore, in high-danger situations, such as if an unusually lethal new variant were identified in one country, it may be preferable to implement preventative measures even prior to observing a changing relationship between indicators and severe outcomes. Mortality is a lagging indicator, following rises in cases and hospitalizations, and changes in transmission dynamics are influenced by other factors (e.g., seasonality, new variants) that have proven difficult to predict (31). As a result, metrics based on mortality should not be construed as leading indicators of future surges, but rather a 'fire alarm' once a surge has begun. However, metrics could be refined to upweight performance during critical periods such as the start of a surge. Finally, future work could also expand these methods to other contexts, such as prediction of combined respiratory disease outcomes (including influenza and RSV). Overall, adaptive metrics may be a powerful tool for designing trustworthy, transparent metrics to guide infectious disease policy.

Data, Materials, and Software Availability. Anonymized cleaned data and code have been deposited in GitHub (https://github.com/abilinski/AdaptiveRiskMetrics). Previously published data were used for this work (public data, URLs in text and on GitHub (32)).

ACKNOWLEDGMENTS. We were supported by the Centers for Disease Control and Prevention through the Council of State and Territorial Epidemiologists (NU38OT000297-02; A.M.B., J.A.S.) and the National Institute on Drug Abuse (3R37DA01561217S1; J.A.S.). Funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Author affiliations: [a]Departments of Health Services, Policy and Practice & Biostatistics, Brown University, Providence, RI 02912; [b]Department of Health Policy, Stanford University, Stanford, CA 94305; and [c]Department of Health Care Policy, Harvard Medical School, Boston, MA 02115

1. CDC, COVID data tracker (2020). https://covid.cdc.gov/covid-data-tracker. Accessed 9 November 2022.
2. CDC, Science brief: Indicators for monitoring COVID-19 community levels and making public health recommendations (2022). https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/indicators-monitoring-community-levels.html. Accessed 9 November 2022.
3. A. Reinhart *et al*., An open repository of real-time COVID-19 indicators. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2111452118 (2021).
4. L. Camera, School Reopening Thresholds Vary Widely Across the Country. US News & World Report. www.usnews.com/news/education-news/articles/2020-08-13/school-reopening-thresholds-vary-widely-across-the-country. Accessed 9 November 2022.
5. E. Shapiro, D. Rubinstein, *Did It Hit 3%? Why Parents and Teachers Are Fixated on One Number* (New York Times, 2020).
6. S. J. Fox *et al*., Real-time pandemic surveillance using hospital admissions and mobility data. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2111870119 (2022).
7. J. A. Salomon, A. Bilinski, Evaluating the performance of Centers for Disease Control and prevention COVID-19 community levels as leading indicators of COVID-19 mortality. *Ann. Int. Med.* **175**, 1240–1249 (2022).
8. World Health Organization, "Regional office for the Western Pacific, calibrating long-term non-pharmaceutical interventions for COVID-19: Principles and facilitation tools" (WHO Regional Office for the Western Pacific, Technical Report WPR/DSE/2020/018, 2020).
9. J. G. Allen, H. Jenkins, *Opinion–The Hard Covid-19 Questions We're Not Asking* (New York Times, 2021).
10. J. K. Varma, *Opinion–When Do Masks Come Off? The Hard Truth About Lifting Covid Restrictions* (New York Times, 2022).
11. B. Rader, Use of At-Home COVID-19 Tests–United States, August 23, 2021–March 12, 2022. *Morb. Mortal. Wkly. Rep. (MMWR)* **71**, 489–494 (2022).
12. D. McPhillips, Covid-19 data reporting is becoming less frequent, making trends harder to track. CNN. https://www.cnn.com/2022/04/25/health/states-scale-back-covid-data-reporting/index.html. Accessed 9 November 2022.
13. J. Adams, Opinion–No, the pandemic 'goal posts' aren't being moved. Wash. Post. The Washington Post. https://www.washingtonpost.com/opinions/2022/01/09/pandemic-goalposts-vaccinations-guidance-jerome-adams-surgeon-general/. Accessed 9 November 2022.
14. E. J. Emanuel, M. Osterholm, C. R. Gounder, A national strategy for the "New Normal" of life With COVID. *JAMA* **327**, 211–212 (2022).
15. Coronavirus (Covid-19) data in the United States (2022). https://github.com/nytimes/covid-19-data. Accessed 20 October 2022.
16. COVID-19 reported patient impact and hospital capacity by state timeseries (2022). https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/g62h-syeh. Accessed 20 October 2022.
17. COVID-19 reported patient impact and hospital capacity by facility (2022). https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u. Accessed 9 November 2022.
18. D. M. Makuc, B. Haglund, D. D. Ingram, J. C. Kleinman, J. J. Feldman, Health service areas for the United States. *Vital Health Stat. Ser. 2, Data Eval. Methods Res.* **112**, 1–102 (1991).
19. J. Berger, *Statistical Decision Theory: Foundations, Concepts, and Methods* (Springer Science & Business Media, 2013).
20. R. Yaesoubi *et al*., Adaptive policies to balance health benefits and economic costs of physical distancing interventions during the COVID-19 pandemic. *Med. Decis. Making* **41**, 386–392 (2021).
21. R. Yaesoubi *et al*., Simple decision rules to predict local surges in COVID-19 hospitalizations during the winter and spring of 2022. medRxiv [Preprint] (2021). https://doi.org/10.1101/2021.12.13.21267657.
22. C. Castillo-Laborde *et al*., Assessment of event-triggered policies of nonpharmaceutical interventions based on epidemiological indicators. *J. Math. Biol.* **83**, 42 (2021).
23. A. Lavazza, M. Farina, The role of experts in the Covid-19 pandemic and the limits of their epistemic authority in democracy. *Front. Public Health* **8**, 356 (2020).
24. D. J. McDonald *et al*., Can auxiliary indicators improve COVID-19 forecasting and hotspot prediction. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2111453118 (2021).
25. E. L. Ray *et al*., Comparing trained and untrained probabilistic ensemble forecasts of COVID-19 cases and deaths in the United States. *Int. J. Forecast.*, https://doi.org/10.1016/j.ijforecast.2022.06.005 (2022).
26. L. M. Stolerman *et al*., Using digital traces to build prospective and real-time county-level early warning systems to anticipate COVID-19 outbreaks in the United States. *Sci. Adv.* **9**, eabq0199 (2023).
27. S. A. Nowak, P. Nascimento de Lima, R. Vardavas, Optimal non-pharmaceutical pandemic response strategies depend critically on time horizons and costs. *Sci. Rep.* **13**, 2416 (2023).
28. P. Nd Lima *et al*., Reopening California: Seeking robust, non-dominated COVID-19 exit strategies. *PLoS ONE* **16**, e0259166 (2021).
29. L. Berger *et al*., Rational policymaking during a pandemic. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2012704118 (2021).
30. N. E. Kogan *et al*., An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Sci. Adv.* **7**, eabd6989 (2021).
31. N. Reich, R. Tibshirani, E. L. Ray, R. Rosenfeld, On the predictability of COVID-19. *Carnegie Mellon University*, 30 September 2021. https://delphi.cmu.edu/blog/2021/09/30/on-the-predictability-of-covid-19/. Accessed 4 April 2023.
32. A. Bilinski, AdaptiveRiskMetrics. GitHub. https://github.com/abilinski/AdaptiveRiskMetrics/tree/main/0_Data/Raw. Deposited 14 February 2023.

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 Information Text

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

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

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, triggering a response 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 the following relationship between the probability of a high outcome, $Pr(Y_{i,w+3})$, and a synthetic hospitalization indicator, $X_{H,i,w}$:

$$logit(Pr(Y_{i,w+3} = 1)) = \beta_0 + \beta_1 X_{H,i,w}, \qquad [1]$$

We then draw $Y_{i,w+3}$ from a binomial distribution with the corresponding probability. We vary simulations across 2 dimensions:

1. **Relationship between inputs and outputs:** The optimal cutoff for a metric with neutral weighting is $\frac{logit(0.5) - \beta_0}{\beta_1} = \frac{-\beta_0}{\beta_1}$. We vary this over time as displayed in Figure S13:
   (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 to 15 and then decreases
2. **Prevalence of "high" outcomes:** 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 Q3 2022.
   (b) Constant: We draw $X_{H,i,w}$ from a $Unif(2, 20)$ distribution for state-times from Q3 2021 through Q3 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 Q3 2022.

For illustration, we set $\beta_0 = -3c$ and $\beta_1 = 3$ 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 S14.

D. Comparison to CDC Community Levels Performance. In the published evaluation of the performance of CDC's Community Levels, the risk designation is considered a true positive if, for two counties with different Community Levels, the county with the higher level had the more severe outcome 3 weeks later (1). This values the ordering of two areas' outcomes as equally important everywhere, whereas our methods prioritize correctly classifying areas on either side of the a priori-specified threshold (e.g., 1 death/100k/week).

Although these are not directly comparable to our measures, we compared our results to the CDC's own published evaluation of the performance of Community Levels (1). These analyses use data from 3/1/2021 to 1/24/2022 to compute the area under the receiver operator curve (AUROC) for Community Levels (comparing High versus Medium/Low) for mortality at 0.71. During approximately the same period (Q2-Q4 2021, which is our training period), we found that Community Levels had unweighted accuracy of 0.67 for predicting >1 death/100k/week and 0.80 for >2 deaths/100k/week at the county level, falling on either side of the AUROC value. We obtained similar estimates for accuracy compared to AUROC estimates of bed occupancy (0.86 CDC AUROC vs. 0.84 accuracy for predicting >10%), but lower estimates for ICU admissions (0.82 CDC AUROC vs. 0.66 accuracy for predicting >2 ICU hospitalizations/100k/week), which may reflect our choice of a lower cutoff as most policy-relevant, but with stronger Community Levels performance at higher cutoff values.

Table S1. Optimal cutoffs for static metrics. Static metrics designated an area as high-risk if all included indicators exceeded their respective optimal thresholds from the training period (4/1/2021-12/31/2021). This table lists optimal cutoffs for each level of geography, risk preference, outcome, and indicator combination. Cutoffs are listed in the order CHO, i.e. 0 5 0 means a cutoff of greater than or equal to 0 per 100K for cases, 5 per 100K for hospitalizations, and 0% inpatient bed occupancy. When an indicator is excluded from a particular functional form, it takes the cutoff 0 by default.

Geography	Risk preference	Outcome	Indicators	Optimal training cutoff
States	Neutral	>1 death/100K/wk	H	0 5 0
States	Neutral	>1 death/100K/wk	HO	0 5 5
States	Neutral	>1 death/100K/wk	C	50 0 0
States	Neutral	>1 death/100K/wk	CH	50 5 0
States	Neutral	>1 death/100K/wk	CHO	50 5 5
States	Neutral	>2 deaths/100K/wk	H	0 15 0
States	Neutral	>2 deaths/100K/wk	HO	0 15 5
States	Neutral	>2 deaths/100K/wk	C	200 0 0
States	Neutral	>2 deaths/100K/wk	CH	200 5 0
States	Neutral	>2 deaths/100K/wk	CHO	200 5 5
States	Don't cry wolf (0.5x FN)	>1 death/100K/wk	H	0 10 0
States	Don't cry wolf (0.5x FN)	>1 death/100K/wk	HO	0 5 5
States	Don't cry wolf (0.5x FN)	>1 death/100K/wk	C	100 0 0
States	Don't cry wolf (0.5x FN)	>1 death/100K/wk	CH	100 5 0
States	Don't cry wolf (0.5x FN)	>1 death/100K/wk	CHO	50 5 5
States	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	H	0 15 0
States	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	HO	0 15 5
States	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	CH	150 15 0
States	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	C	200 0 0
States	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	CHO	200 10 5
States	Better safe than sorry (0.5x FP)	>1 death/100K/wk	H	0 5 0
States	Better safe than sorry (0.5x FP)	>1 death/100K/wk	HO	0 5 5
States	Better safe than sorry (0.5x FP)	>1 death/100K/wk	C	50 0 0
States	Better safe than sorry (0.5x FP)	>1 death/100K/wk	CH	50 5 0
States	Better safe than sorry (0.5x FP)	>1 death/100K/wk	CHO	50 5 5
States	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	H	0 10 0
States	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	HO	0 10 5
States	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	C	150 0 0
States	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	CH	150 10 0
States	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	CHO	150 5 5
HSAs	Neutral	>1 death/100K/wk	H	0 5 0
HSAs	Neutral	>1 death/100K/wk	HO	0 5 5
HSAs	Neutral	>1 death/100K/wk	C	50 0 0
HSAs	Neutral	>1 death/100K/wk	CH	50 5 0
HSAs	Neutral	>1 death/100K/wk	CHO	50 5 5
HSAs	Neutral	>2 deaths/100K/wk	H	0 15 0
HSAs	Neutral	>2 deaths/100K/wk	HO	0 15 5
HSAs	Neutral	>2 deaths/100K/wk	CH	150 10 0
HSAs	Neutral	>2 deaths/100K/wk	CHO	150 10 5
HSAs	Neutral	>2 deaths/100K/wk	C	200 0 0
HSAs	Don't cry wolf (0.5x FN)	>1 death/100K/wk	H	0 10 0
HSAs	Don't cry wolf (0.5x FN)	>1 death/100K/wk	HO	0 5 5
HSAs	Don't cry wolf (0.5x FN)	>1 death/100K/wk	C	100 0 0
HSAs	Don't cry wolf (0.5x FN)	>1 death/100K/wk	CH	100 5 0
HSAs	Don't cry wolf (0.5x FN)	>1 death/100K/wk	CHO	50 5 5
HSAs	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	H	0 15 0
HSAs	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	HO	0 15 5
HSAs	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	CHO	150 15 5
HSAs	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	CH	200 15 0
HSAs	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	C	250 0 0
HSAs	Better safe than sorry (0.5x FP)	>1 death/100K/wk	H	0 5 0
HSAs	Better safe than sorry (0.5x FP)	>1 death/100K/wk	HO	0 5 5
HSAs	Better safe than sorry (0.5x FP)	>1 death/100K/wk	C	50 0 0
HSAs	Better safe than sorry (0.5x FP)	>1 death/100K/wk	CH	50 5 0
HSAs	Better safe than sorry (0.5x FP)	>1 death/100K/wk	CHO	50 5 5

HSAs	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	H	0 10 0
HSAs	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	HO	0 10 5
HSAs	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	CH	100 10 0
HSAs	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	CHO	100 5 5
HSAs	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	C	150 0 0
Counties	Neutral	>1 death/100K/wk	H	0 5 0
Counties	Neutral	>1 death/100K/wk	HO	0 5 5
Counties	Neutral	>1 death/100K/wk	C	100 0 0
Counties	Neutral	>1 death/100K/wk	CH	50 5 0
Counties	Neutral	>1 death/100K/wk	CHO	50 5 5
Counties	Neutral	>2 deaths/100K/wk	H	0 15 0
Counties	Neutral	>2 deaths/100K/wk	HO	0 15 5
Counties	Neutral	>2 deaths/100K/wk	CH	150 15 0
Counties	Neutral	>2 deaths/100K/wk	CHO	150 15 5
Counties	Neutral	>2 deaths/100K/wk	C	200 0 0
Counties	Don't cry wolf (0.5x FN)	>1 death/100K/wk	H	0 10 0
Counties	Don't cry wolf (0.5x FN)	>1 death/100K/wk	HO	0 5 5
Counties	Don't cry wolf (0.5x FN)	>1 death/100K/wk	CH	100 10 0
Counties	Don't cry wolf (0.5x FN)	>1 death/100K/wk	CHO	100 5 5
Counties	Don't cry wolf (0.5x FN)	>1 death/100K/wk	C	150 0 0
Counties	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	HO	0 15 10
Counties	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	H	0 20 0
Counties	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	CH	200 15 0
Counties	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	CHO	200 15 5
Counties	Don't cry wolf (0.5x FN)	>2 deaths/100K/wk	C	250 0 0
Counties	Better safe than sorry (0.5x FP)	>1 death/100K/wk	H	0 5 0
Counties	Better safe than sorry (0.5x FP)	>1 death/100K/wk	HO	0 5 5
Counties	Better safe than sorry (0.5x FP)	>1 death/100K/wk	C	50 0 0
Counties	Better safe than sorry (0.5x FP)	>1 death/100K/wk	CH	50 5 0
Counties	Better safe than sorry (0.5x FP)	>1 death/100K/wk	CHO	50 5 5
Counties	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	H	0 10 0
Counties	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	HO	0 10 5
Counties	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	CH	100 10 0
Counties	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	CHO	100 10 5
Counties	Better safe than sorry (0.5x FP)	>2 deaths/100K/wk	C	150 0 0

Table S2. Switches between high-risk and non-high risk episodes. We estimated the number of switches between predicted high-risk and non-high risk designations using two definitions: 1) any change from the prior week and 2) a change that lasted at least two weeks (following a previous episode at least two weeks in duration). Across states and counties, we find that best-performing adaptive metrics of those in Figure S4 generally predicted fewer unique episodes than CDC Community Levels, and both often predict fewer episodes were observed, with one exception being the outcome >1 death/100K/week. In this case, adaptive metrics had more episodes, but these were substantially closer than static metrics to the values.

Geography	Outcome	Episode definition	True episodes	Adaptive episodes	Community Level episodes
State	>1 death/100K/wk	1	707	450	370
State	>1 death/100K/wk	2	428	309	327
State	>2 deaths/100K/wk	1	447	226	370
State	>2 deaths/100K/wk	2	304	188	327
State	>2 ICU patients/100K/wk	1	280	319	370
State	>2 ICU patients/100K/wk	2	236	254	327
State	>10% inpatient bed occupancy	1	205	320	370
State	>10% inpatient bed occupancy	2	203	240	327
County	>1 death/100K/wk	1	47,978	30,373	29,451
County	>1 death/100K/wk	2	38,299	22,095	20,811
County	>2 deaths/100K/wk	1	46,346	21,410	29,451
County	>2 deaths/100K/wk	2	36,905	16,016	20,811
County	>2 ICU patients/100K/wk	1	23,124	19,065	29,451
County	>2 ICU patients/100K/wk	2	16,762	14,858	20,811
County	>10% inpatient bed occupancy	1	14,058	23,451	29,451
County	>10% inpatient bed occupancy	2	12,103	16,004	20,811

References

1. CDC, Science Brief: Indicators for Monitoring COVID-19 Community Levels and Making Public Health Recommendations (https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/indicators-monitoring-community-levels.html) (2022).

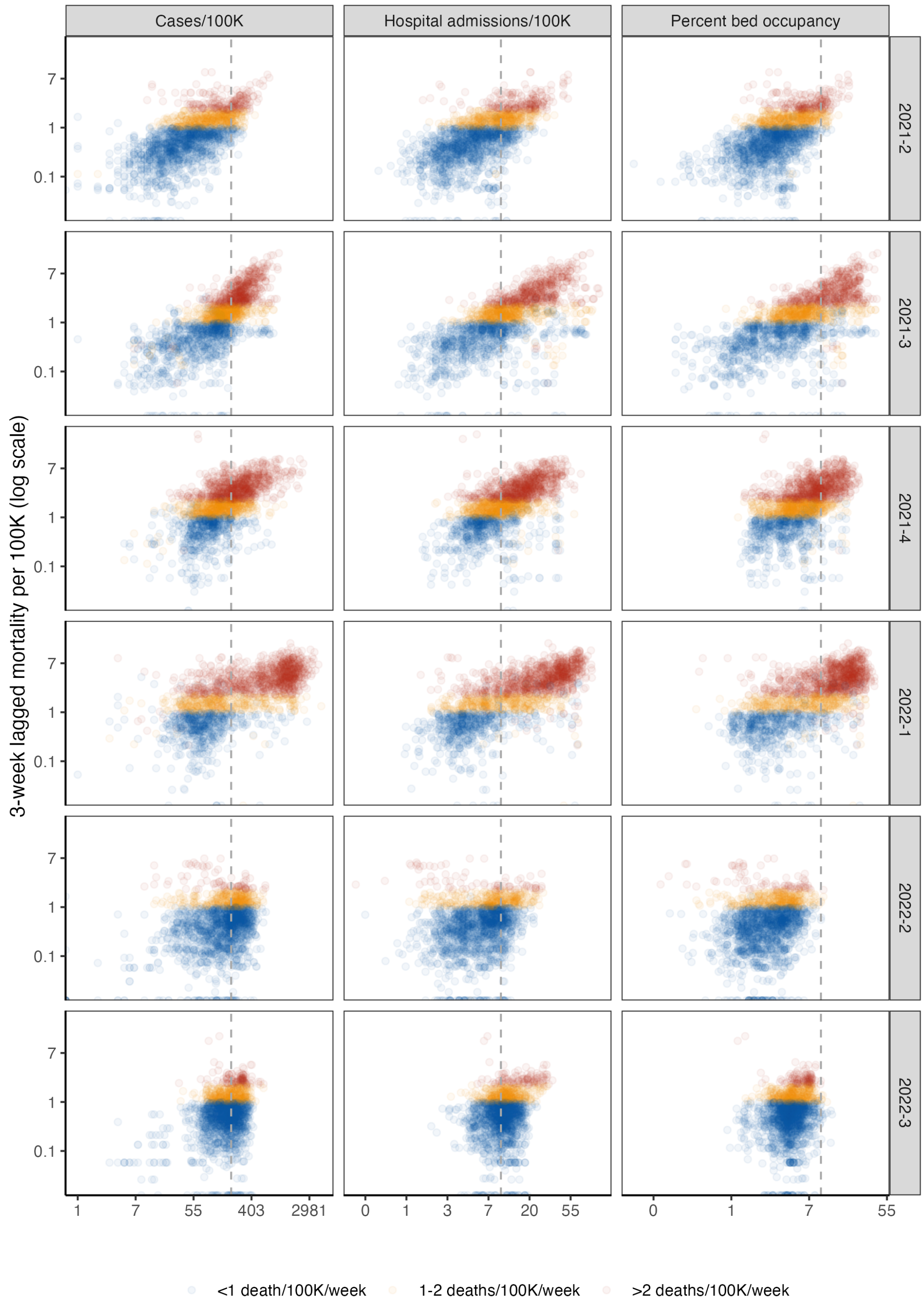

Fig. S1. County-level lagged mortality vs. indicator levels by quarter. Columns indicate different indicators (weekly cases per 100,000 population, new hospital admissions per 100,000 population, and percentage of inpatient beds occupied by COVID-19 patients), and rows indicate quarters. The x-axis displays 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, and counties with greater than 500,000 population are displayed. Colors show mortality outcome level. The vertical gray dotted lines indicate thresholds from CDC Community Levels for each indicator ($\geq$200 cases/100K/week and $\geq$10 new admissions/100K/week or $\geq$10% COVID-19 bed occupancy.)

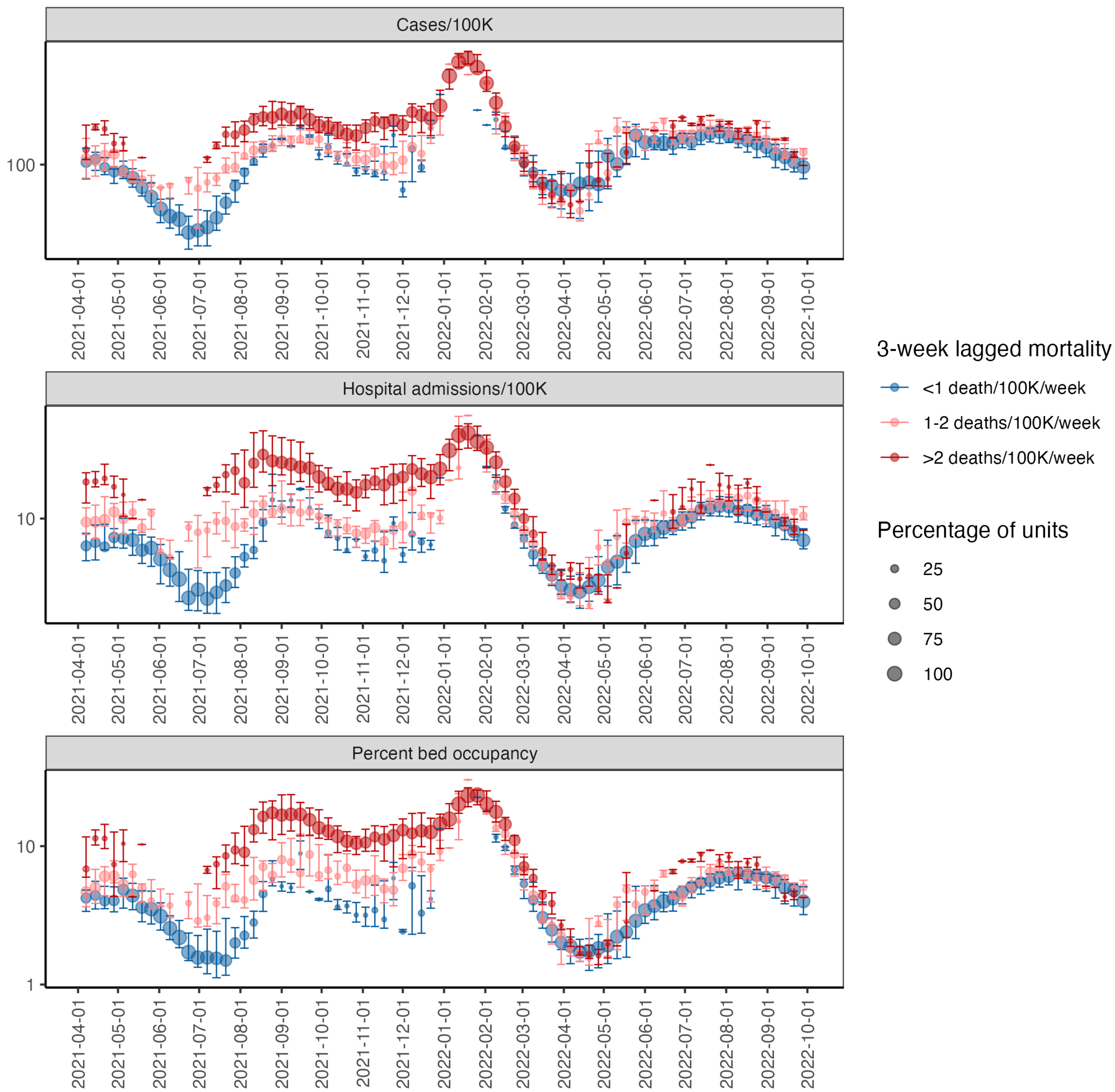

Fig. S2. Indicators by lagged mortality (state). Indicators vary across rows. The x-axis displays time and the y-axis displays the median (point) and interquartile range (bars) of each indicator by future mortality status.

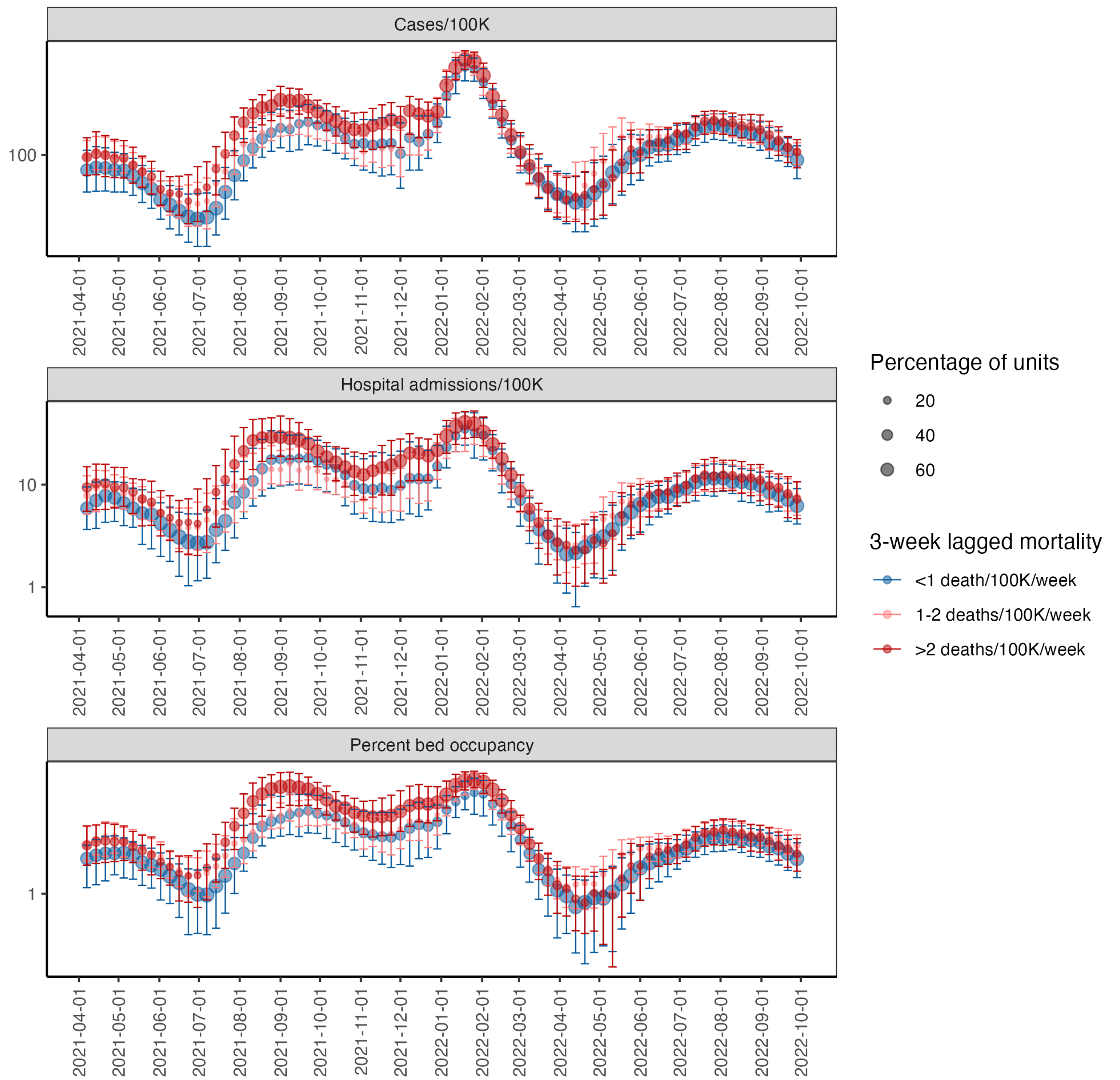

Fig. S3. Indicators by future mortality (county). Indicators vary across rows. The x-axis displays time and the y-axis displays the median (point) and interquartile range (bars) of each indicator by future mortality status.

States

Outcome	Metric	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
>1 death/100K/wk	Adaptive: CHO	88	3	80	10	87	5	86	2	90	3	75	14
	Adaptive: CHOZ	88	3	83	5	88	5	87	2	90	2	80	10
	Adaptive: CHOD	86	6	77	17	85	9	84	4	87	5	74	16
	Adaptive: HZ	89	1	83	5	89	3	87	1	91	3	81	8
	Simplified adaptive: HZ	86	6	82	8	85	12	84	5	89	5	82	4
	Community Levels	64	44	71	24	76	28	72	29	53	62	70	24
	Z	80	15	81	8	81	19	79	12	80	25	83	3
	CHO	86	7	68	41	87	5	63	58	84	11	73	22
	HO	85	7	68	41	87	6	63	58	84	11	73	22
	CH	87	5	56	52	86	9	47	67	90	3	68	26
	H	83	15	56	52	84	19	68	45	88	8	69	26
	C	86	5	45	60	86	9	41	67	90	2	62	33
	Prevalence	68		41		68		41		68		41	
>2 deaths/100K/wk	Adaptive: CHO	87	7	93	3	87	6	94	2	87	8	93	3
	Adaptive: CHOZ	88	6	92	3	88	5	93	3	88	8	92	5
	Adaptive: CHOD	85	10	91	4	86	12	92	5	85	10	90	6
	Adaptive: HZ	90	2	93	3	91	1	93	4	90	2	92	3
	Simplified adaptive: HZ	87	6	94	0	89	3	94	2	88	7	93	1
	Community Levels	88	7	77	36	90	4	72	51	87	12	82	24
	Z	83	11	87	19	85	10	86	24	82	13	89	15
	CHO	89	7	76	37	90	4	75	49	88	6	77	38
	HO	88	6	91	5	90	3	91	10	88	8	81	35
	CH	88	8	70	39	90	2	91	10	88	7	79	35
	H	88	6	91	6	89	5	91	10	87	9	79	38
	C	88	8	70	39	90	4	62	55	88	7	68	42
	Prevalence	36		23		36		23		36		23	

HSAs

Outcome	Metric	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
>1 death/100K/wk	Adaptive: CHO	80	4	73	6	81	3	80	2	84	5	70	11
	Adaptive: CHOZ	83	1	77	1	83	1	81	2	86	1	78	0
	Adaptive: CHOD	80	4	74	7	81	3	81	2	84	5	71	11
	Adaptive: HZ	82	3	77	0	83	1	81	2	86	2	78	0
	Simplified adaptive: HZ	81	3	76	1	80	7	80	3	85	3	78	1
	Community Levels	64	31	68	13	75	14	73	13	52	49	63	27
	Z	78	9	75	2	78	11	74	12	77	16	77	3
	CHO	75	15	70	13	81	4	73	18	68	26	66	16
	HO	75	15	70	13	81	4	73	18	68	26	66	16
	CH	81	4	58	31	81	3	52	47	81	8	66	15
	H	80	6	58	31	81	6	68	26	82	6	67	15
	C	79	5	50	42	80	3	46	54	83	5	64	23
	Prevalence	66		48		66		48		66		48	
>2 deaths/100K/wk	Adaptive: CHO	82	4	88	0	84	5	91	1	83	5	87	2
	Adaptive: CHOZ	84	1	88	0	86	2	91	1	85	1	88	0
	Adaptive: CHOD	81	6	89	0	83	7	90	2	82	5	87	1
	Adaptive: HZ	83	4	88	1	85	3	90	4	84	4	87	1
	Simplified adaptive: HZ	82	5	87	2	83	5	91	1	82	4	86	2
	Community Levels	83	5	76	24	85	2	73	37	81	13	78	14
	Z	80	6	82	11	81	11	81	17	79	9	84	9
	CHO	82	4	80	19	85	3	86	10	82	5	80	17
	HO	82	7	85	5	85	2	86	10	82	7	81	12
	CH	82	3	72	34	85	4	85	12	82	5	75	23
	H	82	7	83	9	84	2	84	15	82	5	75	24
	C	82	7	66	33	85	5	73	27	82	8	66	31
	Prevalence	37		26		37		26		37		26	

Counties

Outcome	Metric	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
>1 death/100K/wk	Adaptive: CHO	75	4	73	4	78	3	79	4	79	4	70	9
	Adaptive: CHOZ	77	0	75	0	79	0	80	2	80	0	75	1
	Adaptive: CHOD	75	3	74	2	77	3	81	1	79	3	71	6
	Adaptive: HZ	75	4	75	1	78	2	80	3	78	5	75	1
	Simplified adaptive: HZ	74	5	74	4	76	5	78	5	79	4	75	1
	Community Levels	67	18	69	8	75	6	73	15	58	35	65	20
	Z	73	6	72	4	74	13	71	12	73	14	73	5
	CHO	74	6	70	9	78	2	73	19	70	16	68	12
	HO	73	7	70	10	77	4	73	21	70	16	68	11
	CH	75	4	57	31	78	2	68	28	78	3	66	13
	H	73	10	56	33	76	4	67	31	78	4	66	13
	C	75	3	54	32	77	2	55	39	78	3	63	21
	Prevalence	58		44		58		44		58		44	
>2 deaths/100K/wk	Adaptive: CHO	79	4	86	1	82	4	88	0	80	4	84	3
	Adaptive: CHOZ	81	1	86	0	84	0	88	1	82	0	85	0
	Adaptive: CHOD	78	5	86	1	81	5	89	0	79	4	85	1
	Adaptive: HZ	80	4	85	2	83	1	87	3	80	4	84	2
	Simplified adaptive: HZ	79	3	85	2	83	3	89	0	79	4	84	2
	Community Levels	80	3	75	23	81	4	72	33	79	9	77	13
	Z	78	6	80	10	78	12	79	15	77	8	81	10
	CHO	80	4	83	5	84	1	85	7	80	5	80	11
	HO	79	4	83	6	83	2	88	0	79	5	80	12
	CH	80	4	82	8	83	1	84	10	80	3	74	21
	H	79	4	82	9	82	3	87	2	79	4	74	23
	C	80	4	65	31	83	3	71	26	79	5	66	28
	Prevalence	35		25		35		25		35		25	

Fig. S4. Head-to-head comparison results, including HSA-level results. The top plots display results from state-level analyses, middle from HSA-level analyses, and bottom from county-level analyses, all 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 positives versus false negatives, with "neutral" corresponding to unweighted accuracy. Prevalence indicates the population-weighted proportion of high location-weeks in a given time period. Secondary outcomes are presented in Figure S5, and weighted accuracy by quarter is presented in Figures S6-S8. For adaptive metrics, models vary functional form to include: 1) CHO (cases, hospitalizations, inpatient bed occupancy); 2) CHOZ (cases, hospitalizations, inpatient bed occupancy, current risk designation); 3) CHOD (cases, hospitalizations, inpatient bed occupancy, weekly changes in each indicator); 4) HZ (hospitalizations, current risk designation); 5) Simplified HZ (hospitalizations, current risk designation – updated quarterly). (For additional adaptive functional forms, see Figure S9.)

States

>2 ICU patients/100K/wk

	Neutral				Don't cry wolf (0.5x FN)				Better safe than sorry (0.5x FP)			
	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR
Adaptive: CHO	87	6	93	5	88	5	92	5	86	7	94	3
Adaptive: CHOZ	86	9	91	7	87	9	89	8	86	7	93	4
Adaptive: CHOD	90	4	93	3	91	0	93	4	90	5	94	3
Adaptive: HZ	86	8	93	5	87	9	92	6	87	7	94	4
Simplified adaptive: HZ	85	14	89	17	86	11	87	20	85	15	91	14
Community Levels	60	42	79	34	73	24	73	48	47	58	86	21
Z	84	12	90	14	85	13	87	19	83	18	92	9
CHO	84	16	61	68	87	9	48	93	80	23	74	44
HO	83	16	61	68	87	9	48	93	80	23	74	44
CH	88	7	38	88	87	6	27	96	89	6	59	57
H	84	17	37	89	81	17	59	77	88	7	58	57
C	87	9	24	90	87	6	22	96	89	3	49	58
Prevalence	73		17		73		17		73		17	

>10% inpatient bed occupancy

	Neutral				Don't cry wolf (0.5x FN)				Better safe than sorry (0.5x FP)			
	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR
Adaptive: CHO	85	5	93	9	85	11	93	9	87	5	95	7
Adaptive: CHOZ	85	6	91	13	85	11	90	16	86	5	94	10
Adaptive: CHOD	88	4	97	0	88	5	96	0	89	2	97	0
Adaptive: HZ	83	10	91	15	84	10	89	22	84	8	94	10
Simplified adaptive: HZ	82	16	93	11	86	9	92	13	80	25	94	11
Community Levels	88	4	74	43	89	6	65	58	87	5	82	29
Z	64	39	89	23	76	23	93	11	52	59	85	37
CHO	87	4	73	44	89	4	78	31	87	4	76	43
HO	85	8	88	12	86	10	91	8	84	10	78	40
CH	88	4	74	43	89	4	77	32	87	4	78	40
H	84	8	88	12	86	10	91	8	83	12	75	44
C	87	4	66	46	89	4	72	32	86	7	78	31
Prevalence	36		11		36		11		36		11	

HSAs

>2 ICU patients/100K/wk

	Neutral				Don't cry wolf (0.5x FN)				Better safe than sorry (0.5x FP)			
	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR
Adaptive: CHO	81	4	91	1	82	4	90	4	83	3	91	3
Adaptive: CHOZ	82	2	91	1	83	2	91	3	83	3	92	1
Adaptive: CHOD	82	3	90	2	83	4	92	1	84	3	91	3
Adaptive: HZ	80	6	91	3	82	3	90	4	82	7	91	3
Simplified adaptive: HZ	80	7	88	11	80	9	86	16	80	13	88	14
Community Levels	66	25	80	25	76	12	74	36	55	39	85	16
Z	80	6	89	9	81	7	86	15	79	13	91	5
CHO	75	14	75	32	81	5	70	43	70	24	82	22
HO	75	14	75	32	81	5	68	45	70	24	82	22
CH	81	6	43	78	83	2	33	95	83	5	62	50
H	80	11	42	78	80	6	62	59	83	4	61	50
C	80	7	42	74	82	3	23	95	83	5	51	53
Prevalence	63		17		63		17		63		17	

>10% inpatient bed occupancy

	Neutral				Don't cry wolf (0.5x FN)				Better safe than sorry (0.5x FP)			
	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR
Adaptive: CHO	84	10	90	15	86	10	90	15	84	6	93	7
Adaptive: CHOZ	84	10	90	15	86	10	90	16	84	7	93	7
Adaptive: CHOD	89	1	95	0	90	1	95	1	88	0	95	0
Adaptive: HZ	83	9	90	15	86	7	87	24	83	8	93	7
Simplified adaptive: HZ	79	18	91	12	84	10	90	15	79	21	93	8
Community Levels	83	8	75	33	82	9	67	45	85	5	83	22
Z	72	26	91	13	81	15	94	4	63	39	88	23
CHO	85	8	84	16	87	4	85	15	84	8	86	17
HO	83	8	91	11	87	6	92	9	82	7	85	19
CH	85	8	77	25	87	5	83	15	84	8	77	29
H	82	10	91	6	86	8	92	6	82	8	90	10
C	85	8	76	26	87	5	79	18	84	8	76	30
Prevalence	28		9		28		9		28		9	

Counties

>2 ICU patients/100K/wk

	Neutral				Don't cry wolf (0.5x FN)				Better safe than sorry (0.5x FP)			
	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR
Adaptive: CHO	81	4	91	2	82	4	91	4	83	4	91	3
Adaptive: CHOZ	82	3	91	2	83	3	91	3	83	3	92	2
Adaptive: CHOD	82	3	91	2	83	3	92	1	84	4	91	3
Adaptive: HZ	81	6	91	3	82	3	91	4	82	6	91	4
Simplified adaptive: HZ	80	7	88	12	80	8	87	15	80	11	88	14
Community Levels	66	25	80	23	76	12	75	34	55	38	86	15
Z	80	6	89	9	81	8	86	14	80	12	91	6
CHO	76	14	76	32	81	4	69	44	70	23	82	21
HO	76	14	76	32	81	4	69	44	70	22	82	21
CH	81	5	45	76	83	3	35	94	83	4	63	48
H	80	11	43	77	81	6	63	58	83	5	62	49
C	79	8	44	71	81	4	25	95	82	7	53	52
Prevalence	63		19		63		19		63		19	

>10% inpatient bed occupancy

	Neutral				Don't cry wolf (0.5x FN)				Better safe than sorry (0.5x FP)			
	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR	Training	Training MR	Test	Test MR
Adaptive: CHO	84	8	90	13	86	8	90	15	84	6	93	6
Adaptive: CHOZ	84	8	90	13	86	8	89	16	84	6	93	6
Adaptive: CHOD	88	1	95	0	89	1	95	1	87	0	95	0
Adaptive: HZ	83	7	90	14	86	5	87	22	83	7	93	7
Simplified adaptive: HZ	79	17	91	11	84	9	90	14	79	20	93	7
Community Levels	84	6	76	31	83	9	68	42	85	4	84	21
Z	72	24	91	13	81	14	94	4	62	38	87	24
CHO	84	7	84	16	87	5	86	13	83	8	86	16
HO	83	7	91	9	87	5	92	7	82	7	91	8
CH	84	7	77	25	87	5	83	15	83	8	77	27
H	82	9	91	5	86	7	92	5	82	7	90	11
C	84	7	76	26	86	5	78	18	83	8	76	28
Prevalence	28		9		28		9		28		9	

Fig. S5. Head-to-head comparison results for secondary outcomes. The top plots display results from state-level analyses, middle from HSA-level analyses, and bottom from county-level analyses, all 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 positives versus false negatives, with "neutral" corresponding to unweighted accuracy. Prevalence indicates the population-weighted proportion of high location-weeks in a given time period. Weighted accuracy by quarter is presented in Figures S6-S8. For adaptive metrics, models vary functional form to include: 1) CHO (cases, hospitalizations, inpatient bed occupancy); 2) CHOZ (cases, hospitalizations, inpatient bed occupancy, current risk designation); 3) CHOD (cases, hospitalizations, inpatient bed occupancy, weekly changes in each indicator); 4) HZ (hospitalizations, current risk designation); 5) Simplified HZ (hospitalizations, current risk designation – updated quarterly)

States

>1 death/100K/wk — Neutral

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	80	90	93	85	83	73	80	88	84
Adaptive: CHOZ	81	90	92	85	84	80	83	88	85
Adaptive: CHOD	80	87	90	83	82	66	77	86	81
Adaptive: HZ	80	92	94	85	84	81	83	89	86
Simplified adaptive: HZ	75	89	94	82	81	83	82	86	84
Community Levels	71	72	50	75	79	59	71	64	68
Z	72	78	91	82	81	80	81	80	81
CHO	79	90	87	87	75	42	68	86	77
HO	79	90	87	87	75	42	68	85	77
CH	76	93	93	88	48	31	56	87	72
H	66	91	94	90	49	31	56	83	70
C	76	89	93	80	24	29	45	86	65
Prevalence	35	75	94	78	17	29	41	68	55

>1 death/100K/wk — Don't cry wolf (0.5x FN)

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	81	88	92	87	89	83	86	87	87
Adaptive: CHOZ	83	88	91	87	89	84	87	88	87
Adaptive: CHOD	81	84	90	85	87	80	84	85	84
Adaptive: HZ	82	93	92	88	90	84	87	89	88
Simplified adaptive: HZ	73	89	93	84	85	84	84	85	85
Community Levels	81	81	66	83	80	55	72	76	74
Z	66	85	91	77	81	80	79	81	80
CHO	80	92	90	89	75	26	63	87	75
HO	79	92	90	89	75	26	63	87	75
CH	81	92	85	86	36	17	47	86	66
H	85	93	75	87	79	39	68	84	76
C	81	92	85	84	23	17	41	86	64
Prevalence	35	75	94	78	17	29	41	68	55

>1 death/100K/wk — Better safe than sorry (0.5x FP)

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	81	93	96	81	78	66	75	90	82
Adaptive: CHOZ	83	93	94	80	82	76	80	90	85
Adaptive: CHOD	80	90	91	80	78	64	74	87	80
Adaptive: HZ	84	92	96	82	82	80	81	91	86
Simplified adaptive: HZ	79	91	96	86	79	80	82	89	85
Community Levels	62	62	34	66	79	64	70	53	61
Z	78	70	91	87	81	80	83	80	81
CHO	79	88	85	85	76	58	73	84	78
HO	79	88	85	85	76	58	73	84	78
CH	81	95	94	88	63	54	68	90	79
H	76	94	96	90	63	54	69	88	79
C	82	93	94	84	49	53	62	90	76
Prevalence	35	75	94	78	17	29	41	68	55

>2 deaths/100K/wk — Neutral

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	95	84	81	87	96	96	93	87	90
Adaptive: CHOZ	96	86	82	87	96	93	92	88	90
Adaptive: CHOD	96	79	80	87	94	92	91	85	88
Adaptive: HZ	95	87	88	87	97	95	93	90	91
Simplified adaptive: HZ	93	83	86	90	97	96	94	87	91
Community Levels	96	88	81	87	85	60	77	88	83
Z	92	81	77	71	95	96	87	83	85
CHO	96	89	81	86	84	59	76	89	83
HO	95	88	82	88	96	91	91	88	90
CH	96	89	80	86	66	57	70	88	79
H	94	88	82	88	95	90	91	88	90
C	96	89	80	86	66	57	70	88	79
Prevalence	5	47	56	61	3	5	23	36	29

>2 deaths/100K/wk — Don't cry wolf (0.5x FN)

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	95	84	83	88	97	98	94	87	91
Adaptive: CHOZ	96	86	83	88	97	95	93	88	91
Adaptive: CHOD	97	78	83	89	96	93	92	86	89
Adaptive: HZ	96	89	88	86	97	95	93	91	92
Simplified adaptive: HZ	94	87	87	89	98	96	94	89	92
Community Levels	96	89	84	89	80	47	72	90	81
Z	91	86	78	66	95	96	86	85	85
CHO	96	90	84	89	87	49	75	90	83
HO	94	89	87	90	95	88	91	90	91
CH	95	89	86	90	95	88	91	90	90
H	92	89	87	90	95	88	91	89	90
C	96	89	84	88	55	43	62	90	76
Prevalence	5	47	56	61	3	5	23	36	29

>2 deaths/100K/wk — Better safe than sorry (0.5x FP)

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	95	86	81	87	95	95	93	87	90
Adaptive: CHOZ	95	88	81	87	95	92	92	88	90
Adaptive: CHOD	96	81	79	87	93	91	90	85	88
Adaptive: HZ	95	87	87	88	95	94	92	90	91
Simplified adaptive: HZ	94	81	88	89	95	96	93	88	91
Community Levels	97	88	77	84	89	73	82	87	85
Z	94	75	77	75	95	97	89	82	86
CHO	94	88	83	86	87	59	77	88	83
HO	89	87	89	90	92	62	81	88	85
CH	94	87	82	86	88	62	79	88	83
H	88	86	88	90	88	59	79	87	83
C	94	88	82	86	64	55	68	88	78
Prevalence	5	47	56	61	3	5	23	36	29

>2 ICU patients/100K/wk — Neutral

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	78	92	90	90	94	94	93	87	90
Adaptive: CHOZ	75	92	90	86	92	94	91	86	88
Adaptive: CHOD	84	95	91	89	96	95	93	90	92
Adaptive: HZ	76	88	95	86	98	94	93	86	89
Simplified adaptive: HZ	80	81	92	74	99	94	89	85	87
Community Levels	68	58	53	91	86	61	79	60	70
Z	76	83	93	77	99	94	90	84	87
CHO	81	79	91	73	82	27	61	84	72
HO	80	79	91	73	82	27	61	83	72
CH	77	92	95	67	41	7	38	88	63
H	67	93	93	65	40	6	37	84	61
C	75	93	94	55	11	5	24	87	56
Prevalence	37	89	92	45	1	5	17	73	45

>2 ICU patients/100K/wk — Don't cry wolf (0.5x FN)

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	81	92	91	87	94	96	92	88	90
Adaptive: CHOZ	77	92	92	82	91	95	89	87	88
Adaptive: CHOD	86	95	93	86	95	96	93	91	92
Adaptive: HZ	77	90	92	82	98	95	92	87	89
Simplified adaptive: HZ	84	84	90	68	99	95	87	86	87
Community Levels	79	72	69	88	81	48	73	73	73
Z	73	88	93	69	99	93	87	85	86
CHO	82	86	93	64	76	3	48	87	67
HO	81	86	93	64	76	3	48	87	67
CH	83	89	88	72	23	0	27	87	57
H	83	85	76	77	80	19	59	81	70
C	83	89	88	68	10	0	22	87	54
Prevalence	37	89	92	45	1	5	17	73	45

>2 ICU patients/100K/wk — Better safe than sorry (0.5x FP)

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	79	92	89	93	95	94	94	86	90
Adaptive: CHOZ	78	92	89	90	94	94	93	86	90
Adaptive: CHOD	83	95	91	91	96	95	94	90	92
Adaptive: HZ	79	88	96	90	97	94	94	87	91
Simplified adaptive: HZ	80	80	95	80	98	94	91	85	88
Community Levels	58	44	38	94	90	74	86	47	66
Z	80	77	93	85	98	94	92	83	88
CHO	79	72	89	82	88	51	74	80	77
HO	79	72	89	82	88	51	74	80	77
CH	81	89	96	78	60	38	59	89	74
H	76	91	95	77	60	38	58	88	73
C	80	92	96	70	41	37	49	89	69
Prevalence	37	89	92	45	1	5	17	73	45

>10% inpatient bed occupancy — Neutral

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	96	77	83	81	98	100	93	85	89
Adaptive: CHOZ	95	76	83	77	98	100	91	85	88
Adaptive: CHOD	95	81	88	90	100	100	97	88	92
Adaptive: HZ	96	76	78	75	99	100	91	83	87
Simplified adaptive: HZ	99	66	81	79	100	100	93	82	87
Community Levels	95	82	86	80	85	57	74	88	81
Z	98	43	51	67	100	100	89	64	77
CHO	95	81	86	80	84	56	73	87	80
HO	93	81	80	79	96	88	88	85	86
CH	95	81	87	80	85	57	74	88	81
H	92	81	80	79	96	88	88	84	86
C	95	81	86	80	66	54	66	87	77
Prevalence	2	57	49	33	0	0	11	36	23

>10% inpatient bed occupancy — Don't cry wolf (0.5x FN)

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	95	74	86	80	98	100	93	85	89
Adaptive: CHOZ	95	74	86	73	97	100	90	85	87
Adaptive: CHOD	94	80	91	89	100	100	96	88	92
Adaptive: HZ	97	75	81	67	100	100	89	84	87
Simplified adaptive: HZ	99	76	82	76	100	100	92	86	89
Community Levels	93	85	87	73	80	42	65	89	77
Z	99	62	68	78	100	100	93	76	84
CHO	98	83	87	78	88	69	78	89	84
HO	98	79	81	81	99	94	91	86	89
CH	98	83	87	77	85	68	77	89	83
H	98	79	81	81	99	94	91	86	89
C	98	83	87	77	73	68	72	89	81
Prevalence	2	57	49	33	0	0	11	36	23

>10% inpatient bed occupancy — Better safe than sorry (0.5x FP)

Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	95	81	83	86	98	100	95	87	91
Adaptive: CHOZ	95	81	83	83	98	100	94	86	90
Adaptive: CHOD	96	84	88	93	100	100	97	89	93
Adaptive: HZ	96	77	80	83	99	100	94	84	89
Simplified adaptive: HZ	98	59	82	82	100	100	94	80	87
Community Levels	96	79	85	86	90	71	82	87	85
Z	97	25	35	56	100	100	85	52	69
CHO	94	80	85	83	88	57	76	87	81
HO	88	81	82	81	93	60	78	84	81
CH	94	80	86	84	89	60	78	87	82
H	86	82	82	81	89	56	75	83	79
C	96	77	85	86	77	69	78	86	82
Prevalence	2	57	49	33	0	0	11	36	23

Fig. S6. State-level results by quarter. Metrics are displayed on the left, with training data from Q2-Q4 2021 and test data from Q1-Q3 2022. Cells report weighted accuracy. Preferences for false positives versus false negatives vary across columns (with "neutral" corresponding to unweighted accuracy) and outcomes across rows. Prevalence indicates the population-weighted proportion of high location-weeks in a given quarter.

HSAs

Neutral

Outcome	Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
>1 death/100K/wk	Adaptive: CHO	72	86	84	82	73	66	73	80	77
	Adaptive: CHOZ	76	87	86	84	74	71	77	83	80
	Adaptive: CHOD	72	85	83	84	74	65	74	80	77
	Adaptive: HZ	76	84	87	84	74	72	77	82	80
	Simplified adaptive: HZ	73	84	86	84	73	72	76	81	78
	Community Levels	63	72	56	71	71	62	68	64	66
	Z	72	78	83	82	72	72	75	78	77
	CHO	70	82	72	78	72	59	70	75	72
	HO	70	82	72	78	72	59	70	75	72
	CH	73	85	83	82	49	41	58	81	69
	H	70	84	84	84	49	41	58	80	69
	C	72	82	84	78	32	39	50	79	64
	Prevalence	44	70	84	79	27	39	48	66	57
>2 deaths/100K/wk	Adaptive: CHO	88	82	76	82	93	90	88	82	85
	Adaptive: CHOZ	89	84	79	82	93	90	88	84	86
	Adaptive: CHOD	88	79	76	82	93	90	89	81	85
	Adaptive: HZ	89	81	80	81	93	90	88	83	86
	Simplified adaptive: HZ	88	80	77	80	92	90	87	82	85
	Community Levels	89	85	75	81	80	66	76	83	79
	Z	83	80	76	71	88	88	82	80	81
	CHO	87	83	76	80	89	71	80	82	81
	HO	89	83	73	78	91	85	85	82	83
	CH	87	83	77	81	78	56	72	82	77
	H	89	83	73	78	90	81	83	82	82
	C	88	84	73	80	62	57	66	82	74
	Prevalence	13	42	56	62	7	10	26	37	32
>2 ICU patients/100K/wk	Adaptive: CHO	79	83	81	84	96	92	91	81	86
	Adaptive: CHOZ	78	84	84	84	96	93	91	82	87
	Adaptive: CHOD	80	85	82	83	96	92	90	82	86
	Adaptive: HZ	78	80	83	82	97	93	91	80	85
	Simplified adaptive: HZ	79	79	83	74	97	92	88	80	84
	Community Levels	74	61	61	85	85	68	80	66	73
	Z	79	80	82	76	97	92	89	80	84
	CHO	77	72	77	76	89	61	75	75	75
	HO	77	72	77	76	89	61	75	75	75
	CH	74	84	85	67	47	15	43	81	62
	H	69	86	85	64	46	15	42	80	61
	C	77	79	83	71	37	19	42	80	61
	Prevalence	31	81	76	40	3	8	17	63	40
>10% inpatient bed occupancy	Adaptive: CHO	97	75	80	72	99	99	90	84	87
	Adaptive: CHOZ	97	75	80	72	99	99	90	84	87
	Adaptive: CHOD	97	85	84	87	99	99	95	89	92
	Adaptive: HZ	97	76	75	72	99	99	90	83	86
	Simplified adaptive: HZ	97	67	75	75	99	99	91	79	85
	Community Levels	92	77	81	74	85	66	75	83	79
	Z	98	59	58	74	99	99	91	72	81
	CHO	96	77	81	76	94	83	84	85	84
	HO	96	77	76	76	99	99	91	83	87
	CH	96	77	81	76	80	74	77	85	81
	H	95	77	74	81	98	94	91	82	87
	C	96	77	81	75	79	73	76	85	80
	Prevalence	2	41	42	26	1	1	9	28	19

Don't cry wolf (0.5x FN)

Outcome	Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
>1 death/100K/wk	Adaptive: CHO	76	84	83	84	82	75	80	81	81
	Adaptive: CHOZ	79	87	83	83	82	77	81	83	82
	Adaptive: CHOD	76	84	84	85	82	75	81	81	81
	Adaptive: HZ	79	86	84	83	83	77	81	83	82
	Simplified adaptive: HZ	72	85	83	84	80	77	80	80	80
	Community Levels	75	81	70	80	75	64	73	75	74
	Z	68	84	83	79	71	72	74	78	76
	CHO	78	85	80	83	78	59	73	81	77
	HO	78	85	80	83	78	59	73	81	77
	CH	77	85	81	83	43	30	52	81	66
	H	79	85	78	82	72	51	68	81	74
	C	76	84	81	82	29	28	46	80	63
	Prevalence	44	70	84	79	27	39	48	66	57
>2 deaths/100K/wk	Adaptive: CHO	92	81	79	83	95	93	91	84	87
	Adaptive: CHOZ	92	84	82	83	95	94	91	86	88
	Adaptive: CHOD	92	79	79	82	95	93	90	83	87
	Adaptive: HZ	92	83	81	80	95	94	90	85	88
	Simplified adaptive: HZ	92	81	77	83	95	93	91	83	87
	Community Levels	90	84	80	84	77	57	73	85	79
	Z	81	85	77	67	88	88	81	81	81
	CHO	91	85	79	83	93	84	86	85	86
	HO	91	84	80	83	93	84	86	85	86
	CH	92	85	78	83	92	82	85	85	85
	H	90	84	80	83	91	79	84	84	84
	C	92	86	77	82	69	67	73	85	79
	Prevalence	13	42	56	62	7	10	26	37	32
>2 ICU patients/100K/wk	Adaptive: CHO	83	83	81	80	97	95	90	82	86
	Adaptive: CHOZ	81	84	84	81	97	94	91	83	87
	Adaptive: CHOD	83	84	81	84	97	95	92	83	87
	Adaptive: HZ	80	82	84	80	98	94	90	82	86
	Simplified adaptive: HZ	82	75	82	68	98	93	86	80	83
	Community Levels	82	73	73	82	81	59	74	76	75
	Z	76	84	83	69	97	92	86	81	84
	CHO	82	79	82	72	87	52	70	81	76
	HO	81	79	82	68	87	50	68	81	75
	CH	81	82	85	66	35	0	33	83	58
	H	82	80	79	75	76	36	62	80	71
	C	80	82	84	62	16	0	23	82	52
	Prevalence	31	81	76	40	3	8	17	63	40
>10% inpatient bed occupancy	Adaptive: CHO	98	75	85	72	99	99	90	86	88
	Adaptive: CHOZ	98	75	84	71	99	99	90	86	88
	Adaptive: CHOD	98	85	87	87	99	99	95	90	93
	Adaptive: HZ	98	78	81	63	99	99	87	86	86
	Simplified adaptive: HZ	98	75	79	72	100	99	90	84	87
	Community Levels	90	76	81	65	80	54	67	82	74
	Z	99	73	72	83	100	99	94	81	88
	CHO	98	81	84	72	95	87	85	87	86
	HO	97	82	81	78	99	99	92	87	89
	CH	98	80	84	73	92	84	83	87	85
	H	97	81	79	81	99	97	92	86	89
	C	97	80	84	71	84	81	79	87	83
	Prevalence	2	41	42	26	1	1	9	28	19

Better safe than sorry (0.5x FP)

Outcome	Metric	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
>1 death/100K/wk	Adaptive: CHO	74	87	89	86	63	62	70	84	77
	Adaptive: CHOZ	79	89	90	89	74	72	78	86	82
	Adaptive: CHOD	74	88	89	86	67	61	71	84	77
	Adaptive: HZ	78	87	91	89	74	72	78	86	82
	Simplified adaptive: HZ	78	86	90	88	74	72	78	85	81
	Community Levels	51	64	42	62	67	61	63	52	58
	Z	76	73	83	86	73	71	77	77	77
	CHO	62	78	65	73	67	60	66	68	67
	HO	62	78	65	73	67	60	66	68	67
	CH	73	88	83	81	59	59	66	81	74
	H	74	88	85	83	59	59	67	82	74
	C	74	87	88	81	51	59	64	83	73
	Prevalence	44	70	84	79	27	39	48	66	57
>2 deaths/100K/wk	Adaptive: CHO	84	85	78	82	91	88	87	83	85
	Adaptive: CHOZ	87	85	82	84	91	88	88	85	86
	Adaptive: CHOD	85	83	78	83	91	88	87	82	85
	Adaptive: HZ	87	81	83	83	91	88	87	84	85
	Simplified adaptive: HZ	86	81	79	83	89	88	86	82	84
	Community Levels	87	85	70	78	83	74	78	81	79
	Z	85	76	76	75	89	88	84	79	82
	CHO	85	83	78	81	87	71	80	82	81
	HO	85	84	76	80	88	76	81	82	81
	CH	85	84	78	80	81	65	75	82	79
	H	83	83	78	80	81	64	75	82	78
	C	85	84	75	80	62	57	66	82	74
	Prevalence	13	42	56	62	7	10	26	37	32
>2 ICU patients/100K/wk	Adaptive: CHO	78	85	87	88	95	90	91	83	87
	Adaptive: CHOZ	80	85	85	88	96	92	92	83	88
	Adaptive: CHOD	78	88	87	87	95	90	91	84	88
	Adaptive: HZ	80	81	84	86	96	92	91	82	87
	Simplified adaptive: HZ	81	75	84	75	96	92	88	80	84
	Community Levels	67	49	50	89	89	77	85	55	70
	Z	81	75	81	84	96	93	91	79	85
	CHO	73	64	72	84	91	71	82	70	76
	HO	73	64	72	83	91	71	82	70	76
	CH	79	83	88	77	65	43	62	83	72
	H	77	85	88	76	64	43	61	83	72
	C	76	86	86	68	46	40	51	83	67
	Prevalence	31	81	76	40	3	8	17	63	40
>10% inpatient bed occupancy	Adaptive: CHO	97	78	79	81	99	99	93	84	89
	Adaptive: CHOZ	96	77	79	81	99	99	93	84	89
	Adaptive: CHOD	97	84	83	88	99	99	95	88	92
	Adaptive: HZ	97	76	77	81	99	99	93	83	88
	Simplified adaptive: HZ	97	63	76	80	99	99	93	79	86
	Community Levels	94	79	81	82	90	77	83	85	84
	Z	97	45	45	65	99	99	88	63	75
	CHO	95	76	80	82	95	82	86	84	85
	HO	90	79	78	79	95	80	85	82	84
	CH	95	76	82	82	78	70	77	84	81
	H	93	78	75	83	97	89	90	82	86
	C	95	76	82	81	77	69	76	84	80
	Prevalence	2	41	42	26	1	1	9	28	19

Fig. S7. HSA-level results by quarter. Metrics are displayed on the left, with training data from Q2-Q4 2021 and test data from Q1-Q3 2022. Cells report weighted accuracy. Preferences for false positives versus false negatives vary across columns (with "neutral" corresponding to unweighted accuracy) and outcomes across rows. Prevalence indicates the population-weighted proportion of high location-weeks in a given quarter.

Counties

>1 death/100K/wk

	Neutral									Don't cry wolf (0.5x FN)									Better safe than sorry (0.5x FP)								
	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	70	79	77	76	75	67	73	75	74	78	78	77	77	83	78	79	78	78	71	83	83	82	65	62	70	79	74
Adaptive: CHOZ	74	80	78	80	76	70	75	77	76	79	81	77	79	83	78	80	79	80	75	83	83	85	73	67	75	80	78
Adaptive: CHOD	71	79	76	80	75	68	74	75	75	78	78	76	81	83	78	81	77	79	72	82	83	82	68	62	71	79	75
Adaptive: HZ	73	76	77	79	76	70	75	75	75	78	79	76	78	84	78	80	78	79	74	78	83	85	74	67	75	78	77
Simplified adaptive: HZ	69	77	76	80	72	69	74	74	74	75	76	75	76	80	79	78	76	77	73	79	83	84	73	68	75	79	77
Community Levels	69	72	60	72	72	63	69	67	68	78	78	71	79	76	64	73	75	74	59	66	48	65	69	63	65	58	62
Z	69	74	76	76	72	68	72	73	73	66	79	76	73	72	68	71	74	72	73	69	76	80	73	68	73	73	73
CHO	72	77	72	77	74	61	70	74	72	79	79	76	80	79	60	73	78	75	67	75	67	73	69	62	68	70	69
HO	72	76	71	77	74	60	70	73	72	77	77	75	80	79	58	73	77	75	66	75	67	74	69	62	68	70	69
CH	70	78	78	80	52	39	57	75	66	79	79	76	80	73	51	68	78	73	72	82	80	79	61	58	66	78	72
H	64	76	77	80	50	37	56	73	64	76	77	75	80	72	48	67	76	71	71	82	81	81	61	57	66	78	72
C	71	79	75	76	44	43	54	75	65	78	80	75	79	45	40	55	77	66	72	81	82	78	53	57	63	78	71
Prevalence	37	62	74	73	25	34	44	58	51	37	62	74	73	25	34	44	58	51	37	62	74	73	25	34	44	58	51

>2 deaths/100K/wk

	Neutral									Don't cry wolf (0.5x FN)									Better safe than sorry (0.5x FP)								
	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	85	79	72	77	91	88	86	79	82	90	79	76	79	94	92	88	82	85	82	83	75	78	88	85	84	80	82
Adaptive: CHOZ	86	81	76	78	91	89	86	81	84	91	83	78	78	94	92	88	84	86	84	83	79	81	89	86	85	82	84
Adaptive: CHOD	86	77	72	78	91	88	86	78	82	90	78	77	79	94	92	89	81	85	82	81	75	80	89	85	85	79	82
Adaptive: HZ	86	78	74	76	91	89	85	80	82	91	82	78	76	94	92	87	83	85	84	79	78	79	89	86	84	80	82
Simplified adaptive: HZ	86	79	73	76	91	89	85	79	82	91	80	78	79	94	92	89	83	86	84	79	75	80	87	86	84	79	82
Community Levels	86	81	73	78	80	66	75	80	77	88	79	76	79	78	59	72	81	77	84	83	70	77	82	73	77	79	78
Z	81	79	73	68	87	85	80	78	79	79	82	73	64	86	85	79	78	78	83	75	73	71	87	85	81	77	79
CHO	87	82	72	76	90	84	83	80	82	90	83	78	79	92	85	85	84	84	83	81	74	77	87	75	80	80	80
HO	86	80	72	76	90	83	83	79	81	90	82	76	79	94	92	88	83	85	82	80	74	78	87	74	80	79	79
CH	87	82	72	76	89	81	82	80	81	90	83	77	79	91	82	84	83	84	82	81	76	78	80	65	74	80	77
H	86	80	72	76	89	80	82	79	80	90	80	75	79	93	90	87	82	84	81	80	75	79	80	63	74	79	76
C	86	81	72	77	62	58	65	80	73	90	83	75	78	69	66	71	83	77	82	81	74	77	63	58	66	79	73
Prevalence	15	39	49	56	9	12	25	35	30	15	39	49	56	9	12	25	35	30	15	39	49	56	9	12	25	35	30

>2 ICU patients/100K/wk

	Neutral									Don't cry wolf (0.5x FN)									Better safe than sorry (0.5x FP)								
	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	80	82	81	84	96	92	91	81	86	83	83	81	80	97	95	91	82	86	78	84	87	88	95	90	91	83	87
Adaptive: CHOZ	78	83	84	84	97	93	91	82	87	81	84	84	81	98	94	91	83	87	80	84	85	88	96	92	92	83	88
Adaptive: CHOD	80	84	82	84	96	92	91	82	86	84	84	82	84	97	95	92	83	88	78	87	87	88	95	90	91	84	87
Adaptive: HZ	79	80	83	83	97	93	91	81	86	81	82	85	80	98	94	91	82	87	80	81	84	86	96	92	91	82	87
Simplified adaptive: HZ	79	79	83	74	97	92	88	80	84	83	76	83	69	98	93	87	80	83	81	76	84	76	96	92	88	80	84
Community Levels	75	61	62	86	85	70	80	66	73	83	73	73	82	82	61	75	76	76	68	49	50	90	89	78	86	55	70
Z	79	80	82	77	97	92	89	80	85	76	84	83	70	97	92	86	81	84	82	75	81	84	96	93	91	80	85
CHO	78	72	77	77	89	61	76	76	76	82	80	82	69	87	51	69	81	75	74	64	72	84	91	72	82	70	76
HO	77	72	78	76	89	61	76	76	76	81	80	82	69	87	51	69	81	75	74	65	73	84	91	72	82	70	76
CH	75	84	85	68	49	17	45	81	63	81	82	84	68	37	1	35	83	59	78	83	87	78	66	45	63	83	73
H	69	86	85	65	47	16	43	80	61	83	80	79	75	76	37	63	81	72	77	86	88	76	65	44	62	83	73
C	77	78	82	71	39	22	44	79	61	80	82	83	62	18	0	25	81	53	75	86	85	70	48	41	53	82	67
Prevalence	32	82	76	41	5	10	19	63	41	32	82	76	41	5	10	19	63	41	32	82	76	41	5	10	19	63	41

>10% inpatient bed occupancy

	Neutral									Don't cry wolf (0.5x FN)									Better safe than sorry (0.5x FP)								
	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall	21-2	21-3	21-4	22-1	22-2	22-3	Test	Training	Overall
Adaptive: CHO	97	75	80	73	99	99	90	84	87	98	76	84	71	99	99	90	86	88	97	77	78	82	99	99	93	84	89
Adaptive: CHOZ	97	75	80	73	99	99	90	84	87	98	76	84	70	99	99	89	86	88	96	77	79	82	99	99	93	84	89
Adaptive: CHOD	97	83	83	86	99	99	95	88	91	98	84	86	86	99	99	95	89	92	97	83	82	88	99	99	95	87	91
Adaptive: HZ	97	76	76	72	99	99	90	83	87	98	79	81	64	99	99	87	86	87	97	76	77	81	99	99	93	83	88
Simplified adaptive: HZ	97	66	75	75	99	99	91	79	85	98	75	79	72	100	99	90	84	87	97	63	77	81	99	99	93	79	86
Community Levels	92	78	81	74	86	68	76	84	80	90	77	81	66	81	57	68	83	75	94	79	80	83	90	78	84	85	84
Z	98	59	59	73	99	99	91	72	81	99	72	72	82	100	99	94	81	87	97	45	45	64	99	99	87	62	75
CHO	96	76	79	76	94	83	84	84	84	97	79	83	73	95	88	86	87	86	95	75	78	82	95	83	86	83	85
HO	96	78	76	77	99	99	91	83	87	97	82	81	79	99	99	92	87	90	94	78	75	84	98	91	91	82	87
CH	96	76	79	76	80	74	77	84	80	97	79	83	73	91	84	83	87	85	95	75	79	81	79	72	77	83	80
H	95	78	74	82	98	94	91	82	87	98	81	79	81	99	97	92	86	89	93	78	75	84	97	88	90	82	86
C	96	76	79	75	79	73	76	84	80	97	79	83	71	83	81	78	86	82	95	75	80	81	77	71	76	83	80
Prevalence	2	41	41	27	1	1	9	28	19	2	41	41	27	1	1	9	28	19	2	41	41	27	1	1	9	28	19

Fig. S8. County-level results by quarter. Metrics are displayed on the left, with training data from Q2-Q4 2021 and test data from Q1-Q3 2022. Cells report weighted accuracy. Preferences for false positives versus false negatives vary across columns (with "neutral" corresponding to unweighted accuracy) and outcomes across rows. Prevalence indicates the population-weighted proportion of high location-weeks in a given quarter.

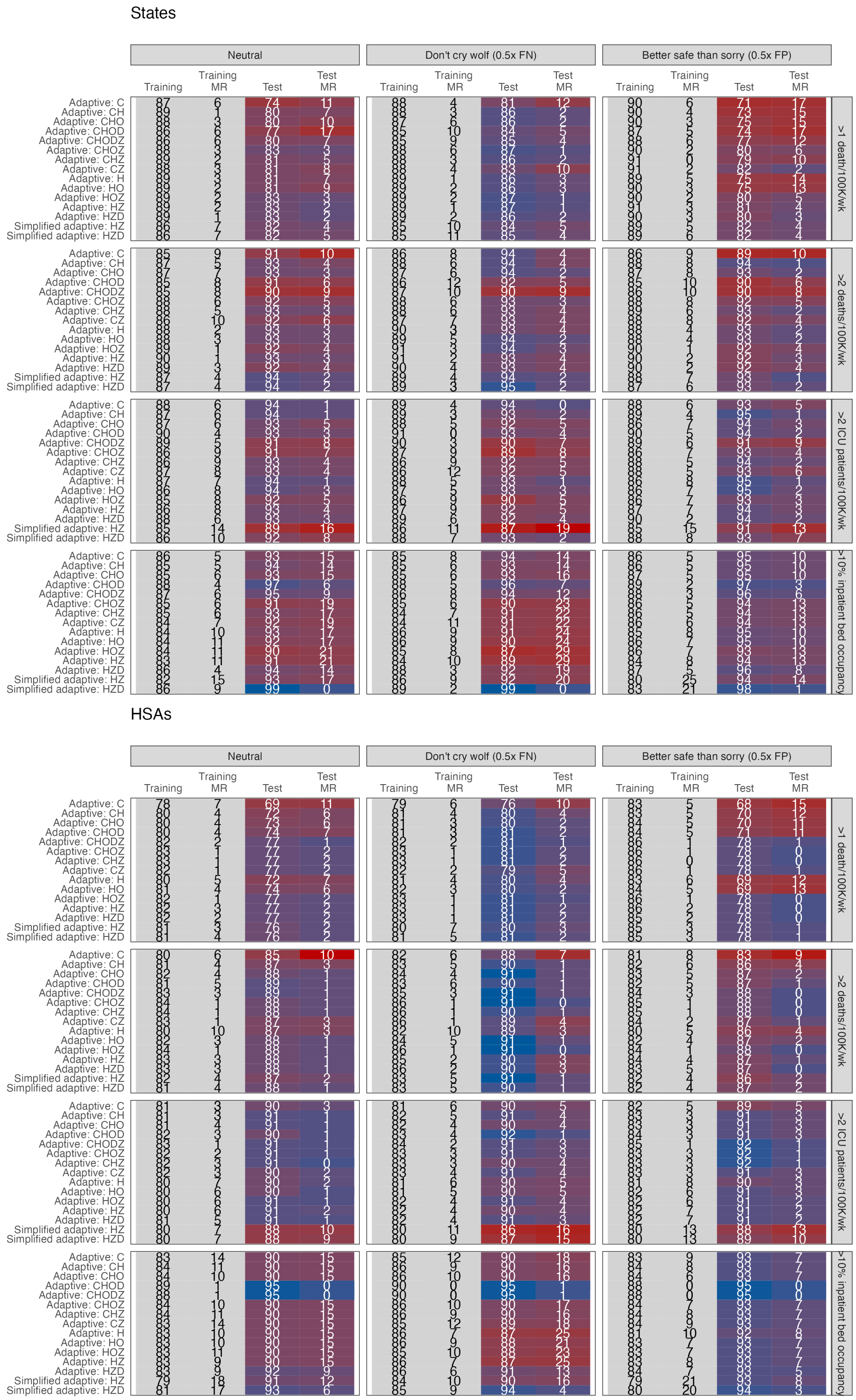

Fig. S9. Head-to-head comparison results with additional adaptive functional forms. The top plots display results from state-level analyses and middle from HSA-level analyses. Metrics are displayed on the left, with training data from Q2-Q4 2021 and test data from Q1-Q3 2022. Cells report weighted accuracy and maximum regret (MR) over training and test periods. Rows vary outcomes, and columns vary preferences for false positives versus false negatives, with "neutral" corresponding to unweighted accuracy. Prevalence indicates the population-weighted proportion of high location-weeks in a given time period. Performance is similar across adaptive specifications, but unusually low-performing specifications (e.g., C and CHOD for >1 death/100K/week) generally had low performance in both training and test periods.

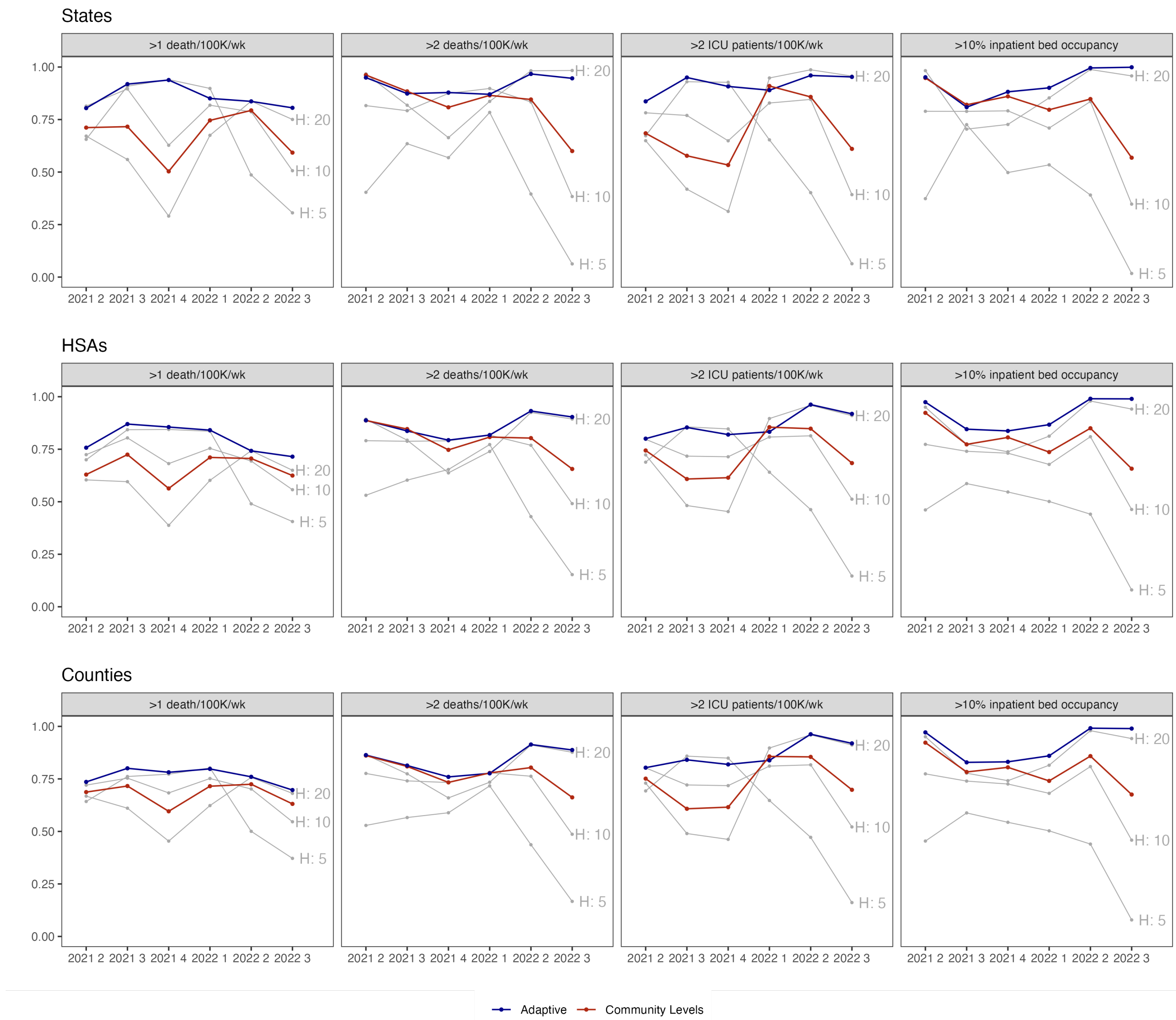

Fig. S10. Weighted accuracy by metric, including HSA-level results. Columns indicate different outcomes. The x-axis indicates quarter, and the y-axis predictive accuracy. Grey lines depict metrics based on new hospital admissions exceeding the labeled threshold. The red line indicates CDC Community Levels and the blue line the best-performing adaptive metric in the pre-intervention period of those listed in Figure S4.

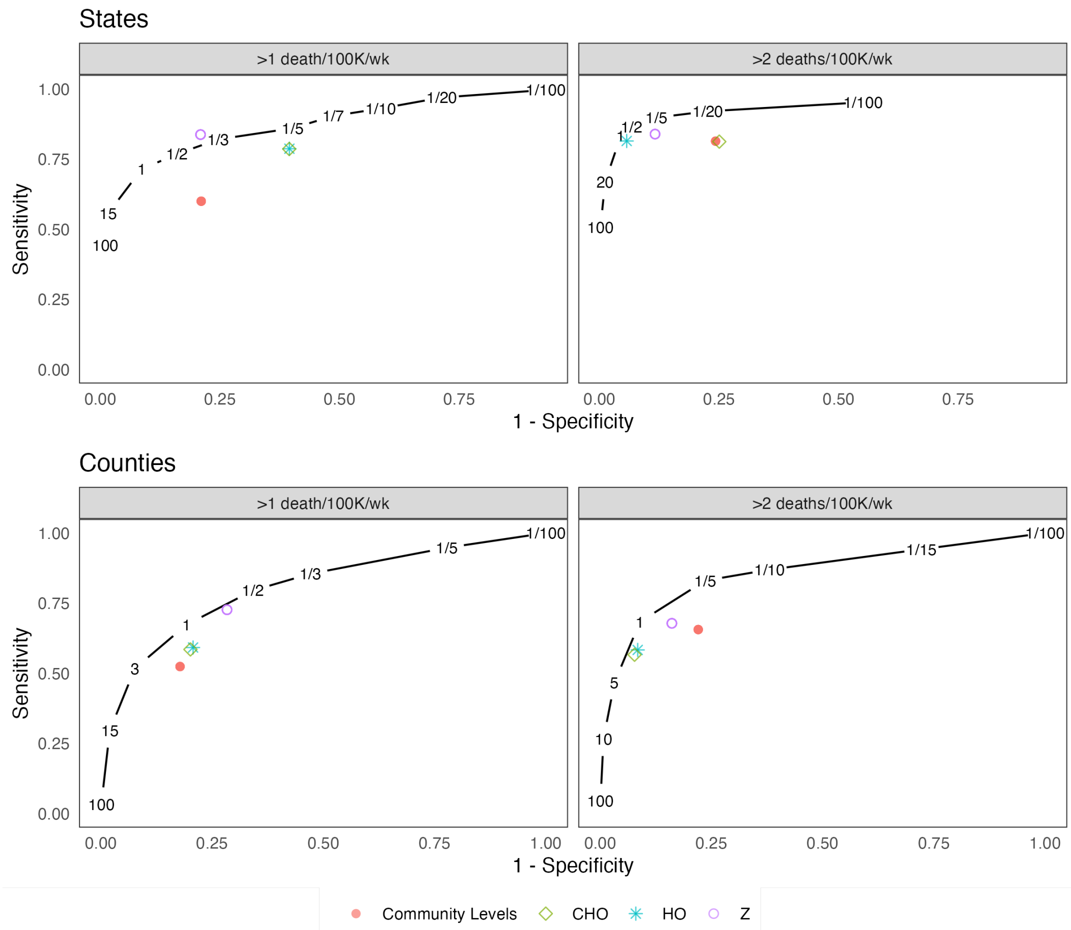

Fig. S11. Receiver operating characteristic (ROC) curves for the test period from January 1, 2022 to September 30, 2022. The black line indicates performance of the adaptive metric (HZ) across different values of wt, indicating the relative preference for false negatives over false positives. The top plot indicates state results and the bottom plot county results.

States

>1 death/100K/wk

	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
Adaptive: CHO	97	4	77	10	97	4	84	2	98	2	70	14
Adaptive: CHOZ	97	4	80	5	97	5	85	2	98	2	76	10
Adaptive: CHOD	90	17	74	17	91	13	82	4	88	20	70	16
Adaptive: HZ	98	3	80	5	98	4	85	1	99	2	77	8
Simplified adaptive: HZ	98	4	79	8	98	3	82	5	99	2	78	4
Community Levels	92	20	66	24	95	12	68	29	90	26	64	24
Z	93	14	78	8	94	11	76	12	91	16	80	3
CHO	98	4	62	41	98	2	57	58	98	4	68	22
HO	98	4	62	41	98	2	57	58	98	4	68	22
CH	99	0	49	46	99	0	37	67	99	0	60	23
H	98	3	48	52	98	4	33	77	99	1	63	26
C	99	0	44	46	99	0	31	67	99	0	58	24
Prevalence	98		31		98		31		98		31	

>2 deaths/100K/wk

	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
Adaptive: CHO	92	3	93	3	90	9	94	2	94	1	92	3
Adaptive: CHOZ	92	3	91	3	91	8	93	3	94	1	90	5
Adaptive: CHOD	92	3	90	4	92	3	92	5	94	0	89	6
Adaptive: HZ	93	0	92	3	93	0	93	4	94	1	92	3
Simplified adaptive: HZ	92	2	94	0	92	2	94	2	93	2	93	1
Community Levels	92	5	73	36	91	9	68	51	93	4	79	24
Z	78	18	88	19	82	13	86	24	73	25	91	15
CHO	93	1	71	49	94	2	91	6	94	0	78	33
HO	93	2	91	5	93	0	91	10	94	0	78	35
CH	93	2	90	6	94	2	91	7	94	1	75	35
H	93	2	90	6	93	0	90	10	93	2	76	38
C	92	6	65	39	91	10	74	27	94	1	63	42
Prevalence	88		9		88		9		88		9	

HSAs

>1 death/100K/wk

	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
Adaptive: CHO	95	2	69	6	94	1	77	2	96	1	65	11
Adaptive: CHOZ	95	2	73	1	93	2	78	2	96	1	75	0
Adaptive: CHOD	95	1	70	7	94	1	78	2	96	1	66	11
Adaptive: HZ	95	0	73	0	94	1	78	2	96	0	75	0
Simplified adaptive: HZ	94	2	72	1	93	3	77	3	96	1	74	1
Community Levels	89	15	63	13	92	6	68	13	86	26	57	27
Z	89	9	72	2	91	6	70	12	87	14	74	3
CHO	92	7	65	13	94	1	69	18	91	15	61	16
HO	92	7	65	13	94	1	69	18	91	15	61	16
CH	94	2	50	31	94	0	44	47	95	3	60	15
H	94	2	51	31	93	1	40	55	95	3	61	15
C	95	2	47	32	94	1	37	54	96	2	58	23
Prevalence	94		39		94		39		94		39	

>2 deaths/100K/wk

	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
Adaptive: CHO	87	3	88	0	85	3	91	1	90	2	86	2
Adaptive: CHOZ	88	2	88	0	86	2	91	1	91	0	87	0
Adaptive: CHOD	87	2	88	0	86	1	91	2	90	1	86	1
Adaptive: HZ	88	0	88	1	86	1	90	4	90	1	86	1
Simplified adaptive: HZ	85	8	87	2	85	5	91	1	87	10	85	2
Community Levels	87	4	73	24	86	3	70	37	88	6	75	14
Z	77	10	83	11	81	5	81	17	73	17	85	9
CHO	87	3	78	21	85	0	86	10	89	3	77	17
HO	87	3	78	21	85	0	86	10	89	3	77	17
CH	87	2	66	40	85	1	84	15	89	4	58	41
H	87	3	65	41	85	1	84	15	89	3	72	24
C	86	7	49	52	85	5	70	27	90	3	54	42
Prevalence	83		14		83		14		83		14	

Counties

>1 death/100K/wk

	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
Adaptive: CHO	89	2	69	4	87	0	77	4	92	1	65	9
Adaptive: CHOZ	89	1	72	0	88	0	78	2	92	1	72	1
Adaptive: CHOD	89	2	71	2	88	0	79	1	92	1	66	6
Adaptive: HZ	89	0	72	1	88	0	78	3	93	0	72	1
Simplified adaptive: HZ	88	4	70	4	87	2	76	5	92	1	71	1
Community Levels	85	10	65	8	86	3	70	15	85	20	60	20
Z	83	6	70	4	85	1	68	12	80	13	71	5
CHO	88	3	66	9	87	0	70	19	89	10	63	12
HO	88	4	66	10	87	0	69	21	89	10	63	11
CH	89	1	52	26	87	0	64	28	91	3	61	13
H	89	2	49	33	87	1	62	31	92	2	61	13
C	89	1	48	32	87	2	48	39	92	1	57	21
Prevalence	88		35		88		35		88		35	

>2 deaths/100K/wk

	Neutral: Training	Neutral: Training MR	Neutral: Test	Neutral: Test MR	Don't cry wolf (0.5x FN): Training	Don't cry wolf (0.5x FN): Training MR	Don't cry wolf (0.5x FN): Test	Don't cry wolf (0.5x FN): Test MR	Better safe than sorry (0.5x FP): Training	Better safe than sorry (0.5x FP): Training MR	Better safe than sorry (0.5x FP): Test	Better safe than sorry (0.5x FP): Test MR
Adaptive: CHO	80	3	86	1	77	4	90	0	85	1	83	3
Adaptive: CHOZ	81	0	86	0	78	1	90	1	85	0	85	0
Adaptive: CHOD	80	2	86	1	78	1	90	0	86	0	84	1
Adaptive: HZ	80	3	86	2	78	3	89	3	85	0	84	2
Simplified adaptive: HZ	78	7	86	2	78	2	90	0	82	9	84	2
Community Levels	80	2	73	23	77	5	71	33	84	3	75	13
Z	71	7	81	10	74	4	79	15	67	14	83	10
CHO	80	1	77	17	79	0	87	6	85	1	76	16
HO	80	1	77	21	78	1	86	9	85	1	76	16
CH	80	1	67	37	79	1	86	8	84	0	72	22
H	80	2	65	40	77	3	83	14	84	1	71	23
C	79	4	62	31	77	5	79	15	84	2	55	38
Prevalence	75		15		75		15		75		15	

Fig. S12. Head-to-head comparison results (omicron training set). The top plots display results from state-level analyses, middle from HSA-level analyses, and bottom from county-level analyses, all 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 positives versus false negatives, with "neutral" corresponding to unweighted accuracy. Prevalence indicates the population-weighted proportion of high location-weeks in a given time period.

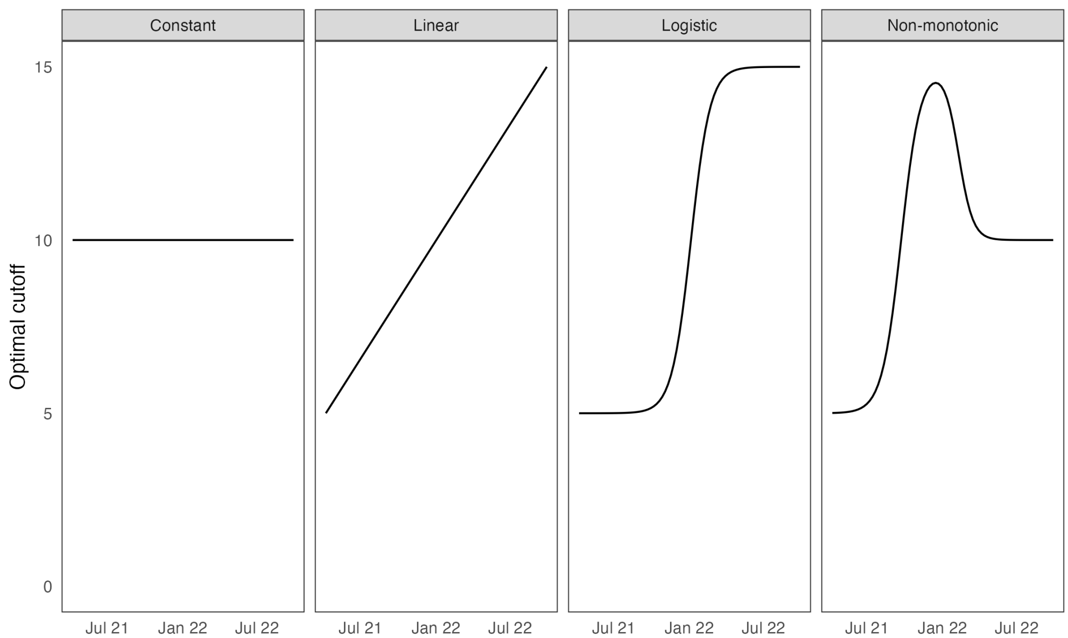

Fig. S13. Simulation scenarios. We vary the optimal cutoff for hospitalization to classify a location-week as "high" over time in different scenarios. The constant scenario assumes a static relationship between indicators and outcomes, while other scenarios assume a changing relationship.

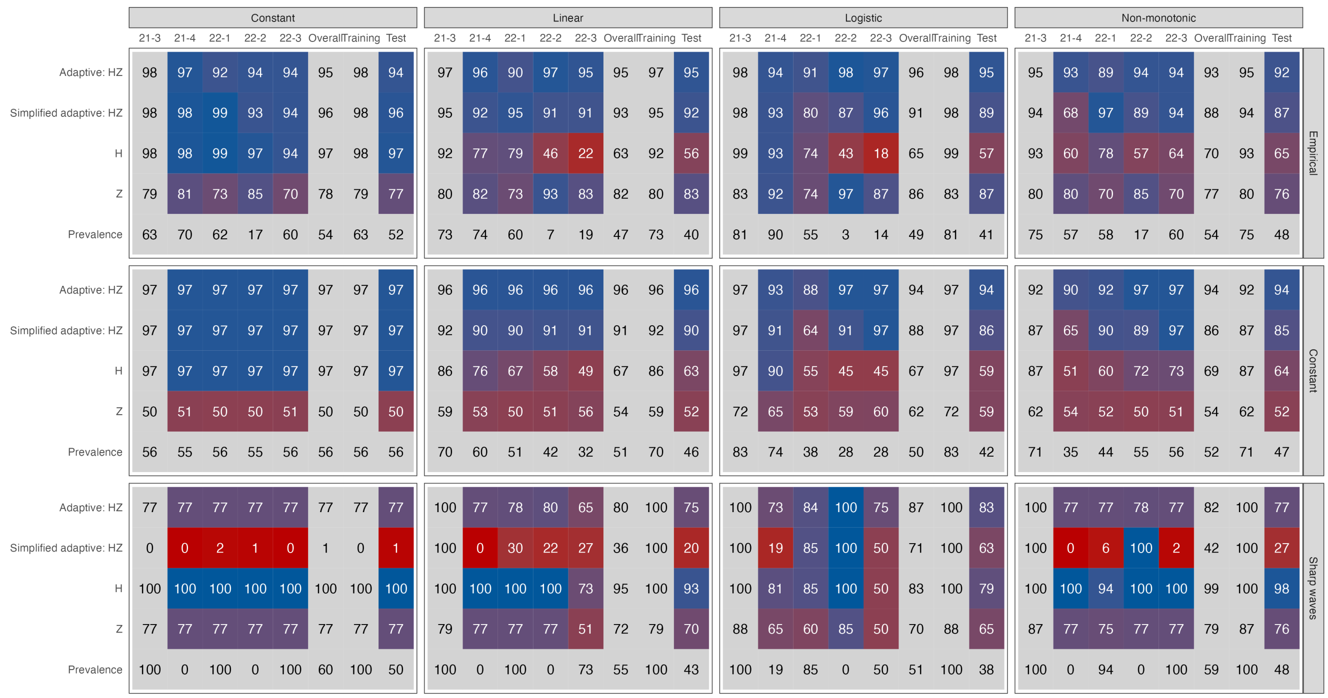

Fig. S14. Simulation results. Columns vary the relationship between the input indicator and outcome over time (Figure S13) and rows vary prevalence of hospitalizations and corresponding high outcomes. Metrics are varied in the y-axis, and 3-week-ahead predictive accuracy is displayed in cells.

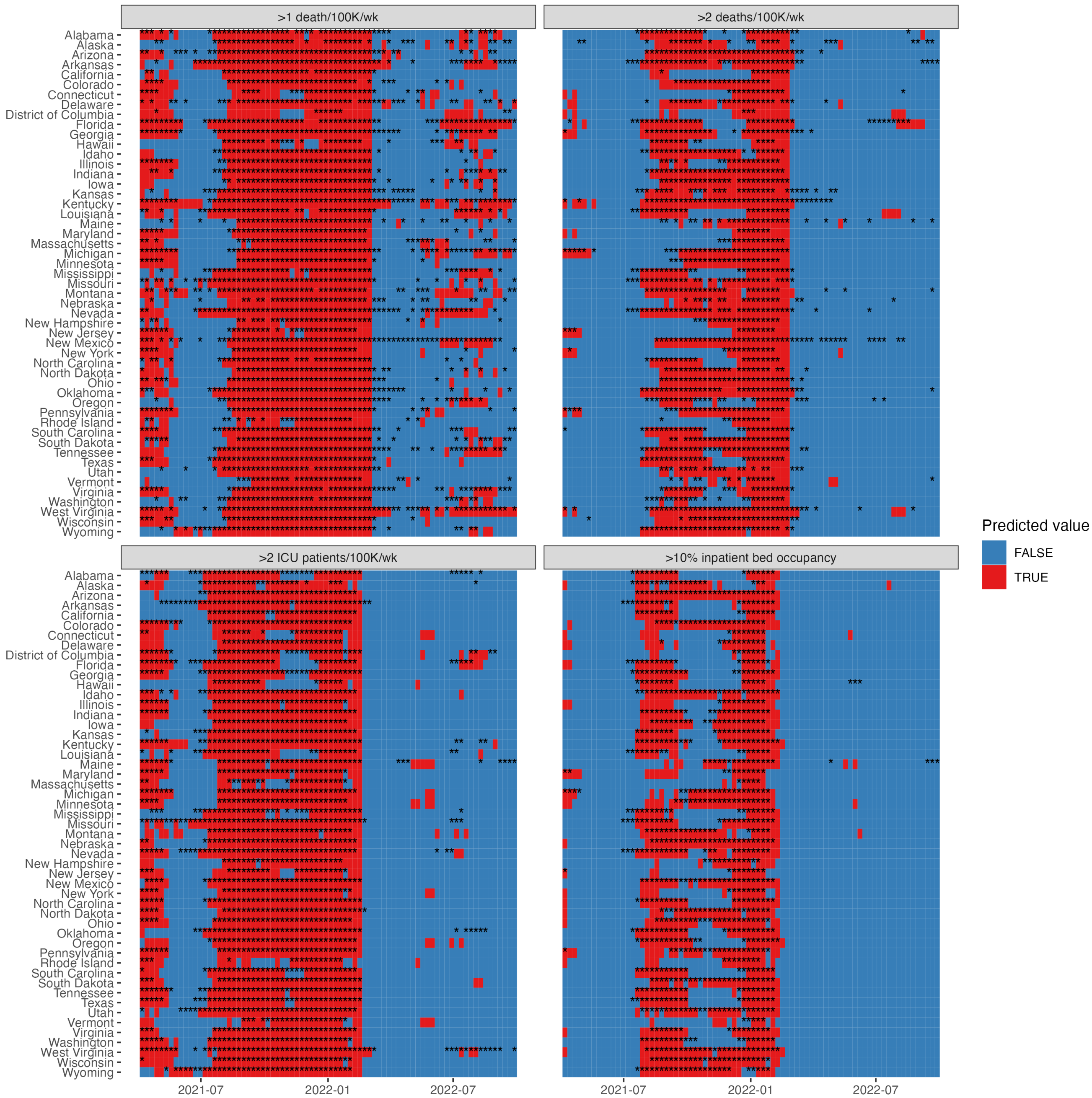

Fig. S15. Predicted and actual values of primary and secondary outcomes at the state level over time. Colors indicate predicted values, while asterisks indicate true values. Predictions are taken from the best-performing adaptive model in the training period, corresponding to those presented in Figure 4.