Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Retrospective Evaluation of Backfill Estimation Models for Influenza, COVID-19, and RSV Hospital Admissions

Michal Ben-Nun

mbennun@predsci.com

Predictive Science Inc.

James Turtle

Predictive Science Inc.

Pete Riley

Predictive Science Inc.

Research Article

Keywords: NHSN hospital admission data, backfill estimation, real-time nowcasting, influenza, COVID-19, RSV

Posted Date: February 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8724860/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Retrospective Evaluation of Backfill Estimation Models for Influenza, COVID-19, and RSV Hospital Admissions

Michal Ben-Nun[1*], James Turtle[1] and Pete Riley[1]

[1]Epi Group, Predictive Science Inc, 9990 Mesa Rim Rd, San Diego, 92121, CA, USA.

*Corresponding author(s). E-mail(s): mbennun@predsci.com;
Contributing authors: jturtle@predsci.com; pete@predsci.com;

Abstract

Hospital admission data for respiratory illnesses such as influenza, COVID-19, and RSV are critical for real-time public health surveillance and short-term forecasting. The National Healthcare Safety Network (NHSN) which provides a comprehensive dataset for all 52 jurisdictions in United States changed its reporting cadence in November 2024. It began releasing preliminary weekly data on Wednesdays and updated versions on Fridays. Because Wednesday reports are often incomplete, especially for the most recent week, users of this dataset must contend with backfill, that is, subsequent (predominantly) increases in reported values as hospitals complete their submissions. In this study, we evaluate four simple nowcasting models to estimate the magnitude of backfill to the last data point: (1) a NULL model that assumes no backfill; (2) a linear-scaling model that assumes backfill scales linearly with the fraction of hospitals yet to report; (3) a prospective-scaling model used operationally during the season, which estimates backfill as a compromise between the two; and (4) a retrospective-scaling model that uses a fitted regression based on the full season's data. Using 26 weeks of data from 52 U.S. jurisdictions, we compared model estimates to the final reported values and calculated absolute and relative errors. The linear-scaling model consistently overestimated backfill and underperformed the NULL model in the majority of cases for all three pathogens. In contrast, the prospective-scaling and retrospective-scaling models outperformed the NULL model for influenza and COVID-19 in most locations, but not for RSV. Despite the assumption that backfill should be non-negative, we observed negative backfill in 21 jurisdictions (often early in the season), which reduced model accuracy. Our results suggest that the intuitive linear-scaling approach is unreliable for this dataset, likely due to heterogeneity in reporting-facility catchment areas. While the retrospective-scaling model yielded the most accurate estimates, its reliance on complete-season data limits operational use. The prospective-scaling model, which requires only current data, provided the most reliable real-time estimates for influenza and COVID-19. These findings support the use of prospective-scaling for operational forecasting and highlight the need for probabilistic nowcasting methods.

Keywords: NHSN hospital admission data, backfill estimation, real-time nowcasting, influenza, COVID-19, RSV

1 Background

The COVID-19 pandemic underscored the importance of high-quality, timely data for modeling and forecasting[1–3]. Since 2020, multiple new data streams have been introduced that provide either

direct (e.g., new hospital admissions) or indirect (e.g., wastewater) measures of respiratory-disease burden. The National Healthcare Safety Network (NHSN, formerly HHS Protect) dataset[4] is considered the gold standard for hospital metrics data related to respiratory diseases including influenza, COVID-19, and RSV[5]. This publicly available dataset provides weekly data on respiratory-related hospital metrics, aggregated at the national and state or territorial levels, as reported to the NHSN—a healthcare-associated infection tracking system run by the Centers for Disease Control and Prevention (CDC). It is available for surveillance weeks beginning in August 2020 and spans changing reporting requirements over multiple time periods.

Throughout the respiratory season, the NHSN dataset is used by modelers, participating in the collaborative Flusight effort[6–9], to produce short-term forecasts (1–4 weeks ahead) of weekly hospital admissions, as well as for more novel categorical forecasting targets (e.g., large increases or decreases). The accuracy of both types of forecasts strongly depends on the value of the most recently reported data point[6, 10, 11]. NHSN admissions data are also used for ongoing real-time efforts to forecast COVID-19[12] and RSV[13], as well as many pandemic-era forecasting contributions[14, 15]. Additionally, the dataset is used by public health and hospital officials for situational awareness, public messaging, and resource allocation[16–18].

Beginning in November 2024, NHSN introduced a preliminary dataset release every Wednesday, and an updated "final" dataset on Friday (more updates can happen and in subsequent releases). This preliminary data has been used weekly by modelers participating in the collaborative FluSight effort. Because hospitals were given only four days to report their latest data (covering admissions up to the Saturday prior to the Wednesday release), it was expected that the data, especially the most recent data point, would be incomplete and subject to backfill in the Friday release and beyond. Although the largest backfill is typically associated with the most recent data point, earlier weeks may also be updated, and in some cases, entire stretches of weeks may change.

Nowcasting refers to the process of estimating the extent of backfill for the most recent (and occasionally earlier) data points. Traditional nowcasting methods[11, 19–24] require access to historical backfill patterns. However, because the new Wednesday–Friday reporting cadence was first implemented during the 2024–25 season, no such historical data existed, limiting the use of these methods. As a participating FluSight team, we developed a simple, operational nowcasting procedure that estimated each Wednesday the expected revision to the latest preliminary value. We then used this backfill-adjusted value as an estimate for the "true" value for model fitting and forecasting.

In this manuscript we describe the operational nowcasting procedure we developed during the 2024–25 season and compare it with three alternatives: (i) a Null model that assumes no backfill, (ii) a linear-scaling model that assumes backfill is proportional to the fraction of hospitals not yet reporting on Wednesday, and (iii) a Retrospective model (not available in real-time) that uses a

fitted regression based on data from the entire season. We evaluate performance using absolute and relative errors, comparing model estimates to both initial and final reports. We perform this analysis for all three respiratory pathogens, but focus on influenza which motivated this work.

The remainder of the manuscript is organized as follows. We begin with a brief discussion of the Data in Section 2 and present the four nowcasting models in Section 3 Methods. The performance of the models is presented in Section 4 Results and discussed in Section 5 Discussion. We conclude with a discussion of the implications of our findings for modelers and public health officials in Section 6 Conclusions.

2 Data

The NHSN dataset is available for surveillance weeks beginning in August 2020 and spans multiple reporting requirements time-periods. The first mandated reporting period for influenza and COVID-19 (requiring daily reporting) began in August 2020 and ended in April 2024, and the current period (requiring weekly reporting) began in November 2024. Data for reporting dates between May 1 and October 31, 2024, represent voluntarily submitted reports. All data and metrics related to respiratory syncytial virus (RSV) were reported voluntarily until November 2024. The weekly reporting cadence, with preliminary data release on Wednesday and a final release two days later, on Friday, began in the current reporting period (November 2024). For the purpose of this study we used 26 weeks of data from 52 U.S. jurisdictions (all 50 states, Puerto Rico, and District of Columbia) which cover the 2024-25 respiratory season (November 2024 to May 2025). We excluded the week of January 18, 2025, when reporting was suspended.

Whereas the NHSN dataset includes 190 columns, only a small subset is used in this analysis: the epidemic week end date, jurisdiction, the total number of new hospital admissions of patients with confirmed influenza, COVID-19, or RSV during the reporting week (three columns), and the proportion of hospitals reporting new admissions of patients with confirmed influenza, COVID-19, or RSV during the same week (three columns). We use these reporting-fraction fields to develop a simple procedure for estimating expected "backfill" for the most recently reported week. During the study period, we first consider all 1,352 location–week combinations (26 weeks $\times$ 52 jurisdictions), but as discussed below we exclude a small number of them when comparing model estimates with the initial and final reported values.

Across the study period, most values were revised only a small number of times, and revisions generally result in increased values, though the extent of decrease in values was greater than we expected and we discuss it in the following sections. Revisions tend to be small, with a median difference of 3.4% (mean 8.1%) between the highest and lowest values observed for a given surveillance

week over the study period. However, a small number of values undergo large revisions, and some states exhibit substantially more revisions than others (e.g., New York, Texas, and Washington).

In Figure 1, we show preliminary (green) and final (red-orange) influenza hospital admission data (solid lines, right y-axis) and reporting fraction data (dashed lines, left y-axis) for the state of Arizona. For each epidemic week, the preliminary curves reflect the data as first reported on the corresponding Wednesday. Focusing first on the left panel, which displays data from the early respiratory season, we observe that the preliminary data point for the week ending on December 14, 2024, could have been interpreted—when first published on December 18, 2024—as either a true sharp decline in new hospital admissions (from 992 to 438), an artifact of a low preliminary reporting fraction (0.67), or some combination of the two. While the subsequent revision of this data point to 1,492 admissions represents an unusually large backfill, this example highlights the challenge faced by modelers, public health officials, and hospital administrators in evaluating the reliability of the most recently reported data point—especially in real time—and, to a lesser extent, the data from preceding weeks.

The right panel of Figure 1 presents the preliminary and finalized data for the full respiratory season in Arizona. Multiple large updates are evident. While all revisions were associated with an increase in the fraction of hospitals reporting, not all backfill adjustments were positive—that is, some data points were revised downward even when the reporting fraction increased. This is particularly noticeable in the long stretch of downward revisions observed early in the season (left panel).

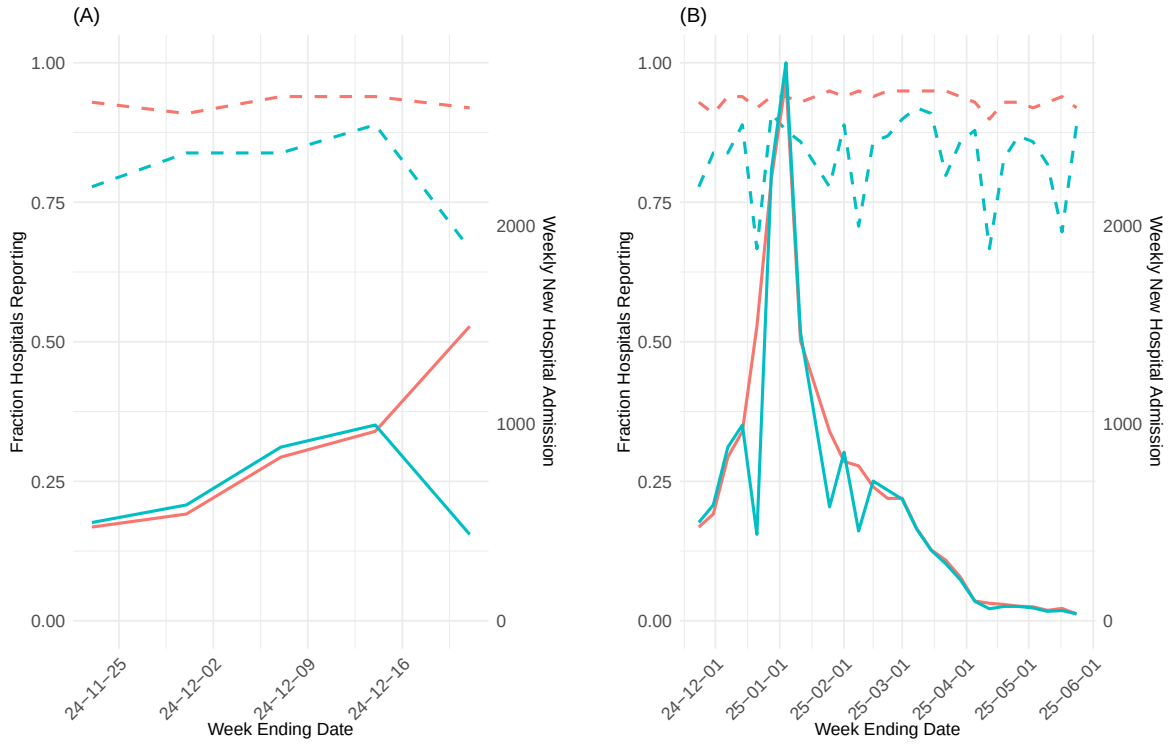


Fig. 1 Initial/preliminary (green) and final/current (red-orange) fractions of hospitals reporting new admissions of patients with confirmed influenza (dashed lines, left y-axis), along with the initial (green) and final (red-orange) total number of new hospital admissions with confirmed influenza (lines, right y-axis). The data shown are for the state of Arizona. The left and right panels display data from the early respiratory season and the full respiratory season, respectively.

3 Methods

As expected and illustrated in Figure 1 for the state of Arizona, preliminary reporting is incomplete. When designing our model to estimate the magnitude of backfill, we made a single simple assumption that this initial reporting fraction would not decrease in subsequent updates and that, consequently, the number of confirmed admissions would not decrease. This was the only assumption we made about data updates.

In practice, when analyzing the most up-to-date data, we found that our assumption—that reporting fractions would only increase—held true for all but 11 of the 1352 location-week combinations considered in the analysis (52 locations x 26 weeks = 1352). Motivated by our model assumption that reporting can only increase during updates, we exclude these 11 location-week combinations from the analysis. We also excluded seven location-week combinations with a very low (below 0.3) initial reporting fraction. In total, we removed 18 location-week combinations out of 1352 (1.3%). Supporting Information Figures A1 and A2 display all the data included in the analysis: initial and final/current hospital admission data and initial and current reporting fraction.

In this retrospective analysis, we compare the performance of the model used during the FluSight challenge season with three alternative linear models. Each Wednesday, when a preliminary report

was published and a forecast was due, we calculated, for each location, the *ratio* between the current reporting fraction and the maximum observed reporting fraction to date (max/current). The most recent data point was then scaled by $(1+ratio)/2$, and this adjusted value was used for model fitting, forecasting, and situational awareness. We refer to this model as prospective-scaling. The operational prospective-scaling is a compromise between two models: (i) A linear scaling model, which assumes that the backfill scales proportionally with the missing fraction of reporting facilities. For example, if only 80% of facilities have reported, the preliminary count is scaled by 1/0.8 to account for expected backfill. This model assumes that all reporting hospitals have similar catchment areas and, under the additional assumptions of similar disease prevalence and age distribution, supports linear scaling of incomplete data. (ii) A NULL model that assumes no backfill—that is, no adjustments are made to the preliminary data. This model is motivated by the idea that introducing corrections may lead to greater error than leaving the data unadjusted. Given that preliminary reporting fractions are generally high, this assumption is not unreasonable. The prospective-scaling model is thus a simple compromise between these two approaches. The third model we compare is a retrospective model, which can only be resolved after the season has ended. It is used both for the analysis and as a potential candidate for future seasons. In this model, a linear regression is performed for each location, using the initial reporting fraction as the predictor and the ratio of final/current to preliminary values as the response. We expect a negative correlation and a negative regression slope, since higher initial reporting fractions should correspond to ratios closer to one.

Example regression fits for six locations are shown in Figure 2, and similar plots for all 52 locations are provided in Figure A3 of the Supporting Information. As expected, most locations show a negative correlation and slope. However, there are exceptions. Some locations, such as Vermont, show little to no correlation (and a slope near zero), while one location, Iowa, exhibits a positive correlation and slope. We discuss possible reasons for these deviations in Section 5.

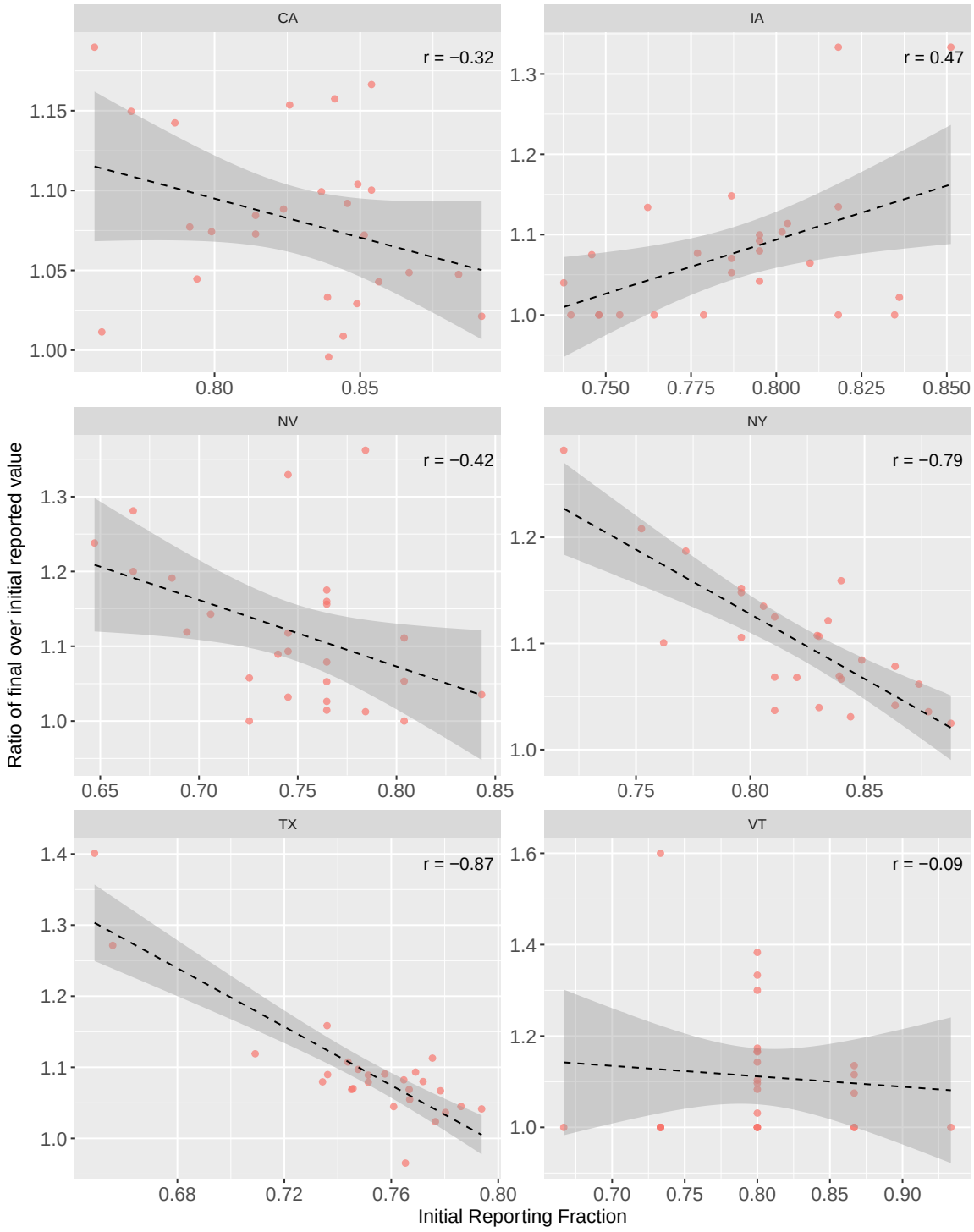


Fig. 2 Currently reported value divided by the initially reported value versus the initial reporting fraction for six representative locations. The legend in each panel indicates the correlation between the two variables; the dashed line represents the linear regression fit, and the shaded area shows the 95% confidence interval.

For each location, we used the regression to reconstruct a retrospective time series, which we refer to as the retrospective-scaling model. We also constructed prospective-scaling, linear-scaling, and NULL time series models, all of which are based on the initially reported values, and compared

each one to the currently reported data for the analysis period. For each location-week, we evaluated three metrics: error, relative error, and error relative to the NULL model. The error is calculated as the difference between the current value and the model estimate. The relative error is obtained by dividing this difference by the current value. The error relative to the NULL model is computed by dividing the model's error by the error of the NULL model for the same location-week. Additionally, we applied the same methodology to the other two respiratory viruses included in the NHSN dataset: COVID-19 and RSV.

4 Results

Figure 3 shows for each week in the study period the currently reported value (black line), linear scaling value (orange line), null model value (red line), prospective-scaling value (blue) and retrospective-scaling (purple) for six representative states. Equivalent plots for all 52 locations are provided in Figure A4 of the Supporting Information.

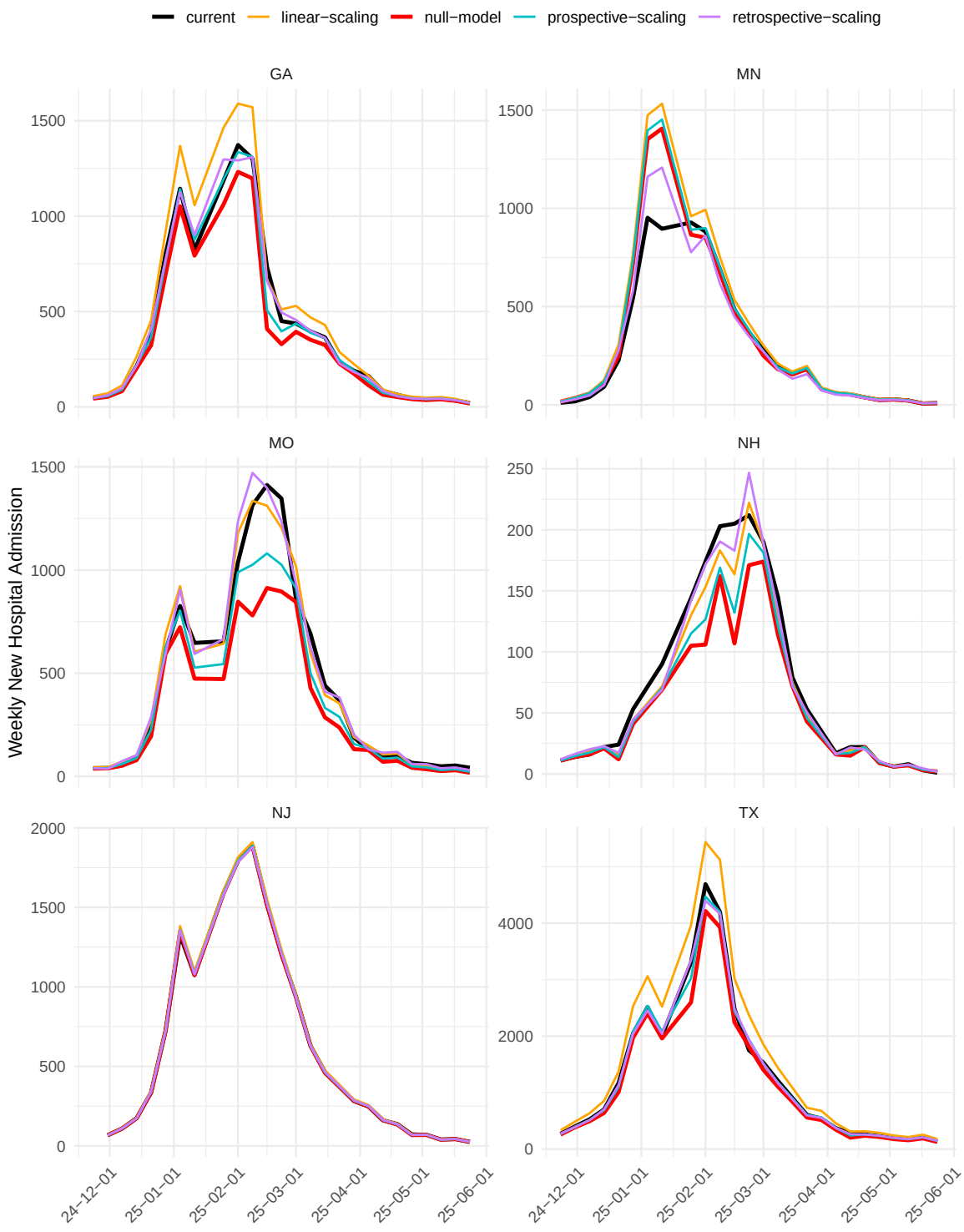


Fig. 3 Reported and models time-series for six representative locations. Black line: current value, orange line:linear-scaling model, red line: NULL model (corresponding to the initial reporting of each value), blue line: prospective-scaling, purple line: retrospective scaling.

As a reminder, the NULL model corresponds to the initial reported values. Therefore, the difference between the current data (black line) and the NULL model (red line) represents the observed backfill. Differences between the other models and the NULL model represent various estimates of

the backfill, and their distance from the current data reflects their error, which we quantify in detail below.

We observe a range of behaviors in both the backfill and model estimates (see also Figure A4 in the Supporting Information). Large backfills are not uncommon in some locations (for example, in Minnesota (MN)), and they are not limited to peak weeks. As expected, most backfills are positive; however, in MN, the first eight weeks of the season exhibit negative backfill, and during some weeks the magnitude is substantial. Table 1 lists the number of weeks and locations with negative backfill. While most locations had no instances of negative backfill (and are therefore not included in the table), a few (four) experienced negative backfill during more than 25% of the 26-week analysis period. For all other locations not shown in Table 1, there were no instances of negative backfill during the 26-week period.

Table 1 The number of weeks and locations with negative backfill. All locations not shown in the table had no instances of negative backfill during the 26-week analysis period.

Number of weeks	Locations
10	AZ, LA, MT
8	MN
5	DC, OH
4	CT, MI, WA
3	NH, PR
2	DE, ID, WY
1	AR, CA, MD, NJ, TX, WV

Figure 4 shows the three error measures for two states: Florida (FL, top row) and Kentucky (KY, bottom row). The errors and relative errors of the linear-scaling, NULL model, prospective-scaling, and retrospective-scaling models are shown in orange, red, blue, and purple, respectively, in the left and middle columns. The right column displays the errors relative to the NULL model for the linear-scaling, prospective-scaling, and retrospective-scaling models, using the same color scheme (excluding red for the NULL model).

The dashed vertical line in each panel marks the peak week. In the left and middle columns, negative and positive values indicate model overestimation and underestimation of the backfill, respectively. In the right column, horizontal dashed lines at +1 and -1 are shown: values within this range indicate that the model error was smaller than the NULL model error. Negative values in this column indicate that the model and NULL model had errors of opposite signs, while positive values indicate errors of the same sign.

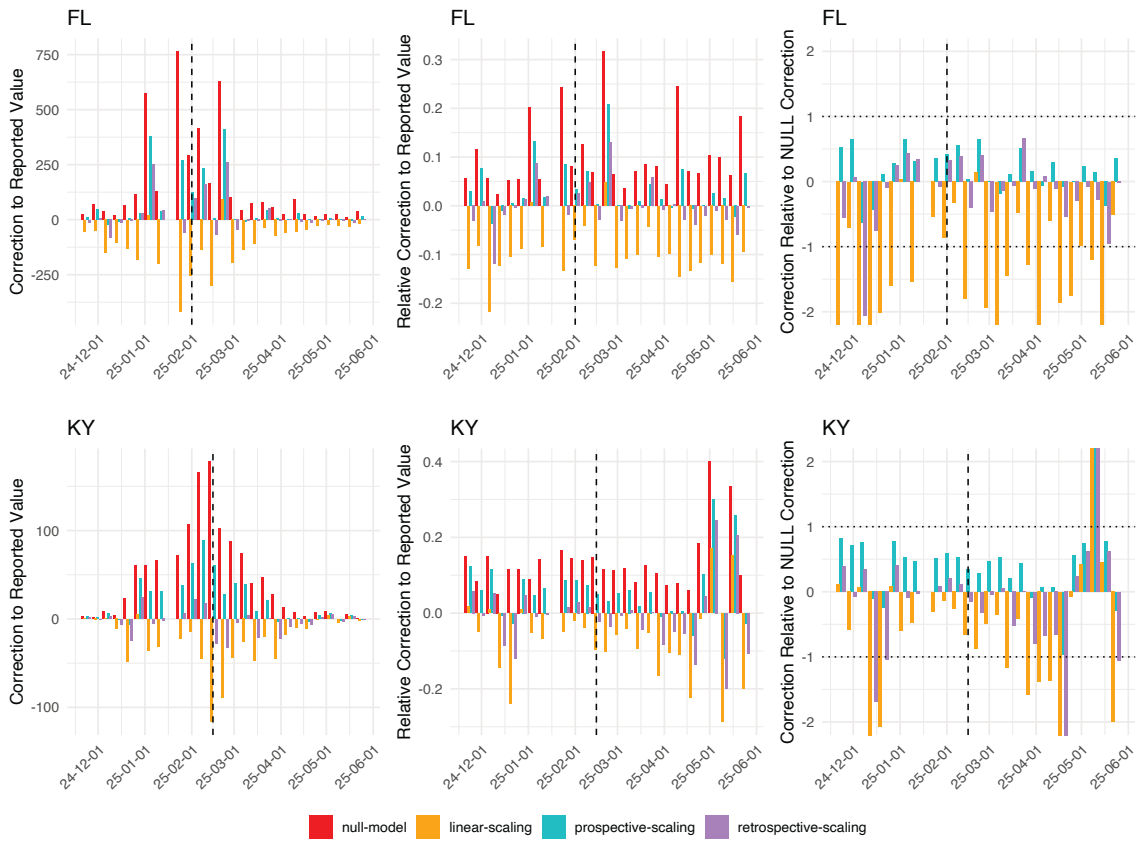


Fig. 4 Error measures for the NULL model (red), linear-scaling model (orange), prospective-scaling model (blue), and retrospective-scaling model (purple) for Florida and Kentucky (top and bottom rows, respectively). Left column: error. Middle column: relative error. Right column: error relative to the NULL model error. The dashed black line in each panel indicates the peak week.

The NULL model errors are all positive in both states because all backfill values were positive, i.e., the final values are always higher than the initial reports. The linear-scaling model overestimated the magnitude of the backfill for all weeks in Kentucky and for all but one week in Florida. The retrospective-scaling model displays significantly lower errors than both the NULL and linear-scaling models, and includes both positive (underestimation) and negative (overestimation) errors.

For these two states, the prospective-scaling model errors are almost always positive, indicating a mostly underestimation of backfill. Errors for all models tend to be largest around the peak week; however, relative errors are not necessarily largest around peak week and are often greater at the beginning or end of the season, when reported values are smaller.

When comparing model errors to the NULL model (right column of Figure 4), we see that in Florida, the linear-scaling model performs worse than the NULL model for most weeks, whereas in Kentucky, it performs better than the NULL model for the majority of weeks. Both the retrospective-scaling and prospective-scaling models perform better than the NULL model for: (i)

All weeks (prospective-scaling in Florida), (ii) All but one week (prospective-scaling in Kentucky and retrospective-scaling in Florida), or (iii) Most weeks (retrospective-scaling in Kentucky).

Figures A5– A7 in the Supporting Information display these three error measures for all 52 locations. Here too, we observe that: (i) for nearly all location-week combinations, the linear scaling model overestimates the extent of the backfill; (ii) the NULL model errors are positive, except when the backfill is negative; and (iii) the prospective- and retrospective-scaling model errors include both positive and negative values (i.e., both over and under estimation of the backfill).

To characterize the performance of the linear-scaling, prospective-scaling, and retrospective-scaling models relative to the NULL model, we calculated—for each model and location—the fraction of weeks in which the model had a lower error than the NULL model. The results are shown in Figure 5, which extends the analysis to the other two respiratory pathogens included in the NHSN dataset: COVID-19 and RSV. A value of 1 (blue) indicates that the model outperformed the NULL model for every week in the analysis period for a given location, while a value of 0 (brown) indicates that the NULL model performed better in every week. Values greater/lower than 0.5 indicate that the model outperformed the NULL model for more/less than half of the 26 weeks included in the analysis period. This figure, along with Table 2, highlights the poor performance of the linear-scaling model relative to the NULL model across all three pathogens. For influenza, the linear-scaling model outperformed the NULL model for more than half of the season in only 8 out of 52 locations. The corresponding numbers for COVID-19 and RSV are 7 and 8, respectively. In contrast, the prospective-scaling model performed better than the NULL model for more than half of the season in 32, 37, and 24 locations for influenza, COVID-19, and RSV, respectively. Similarly, the retrospective-scaling model outperformed the NULL model in 33, 38, and 23 locations for the same three pathogens.

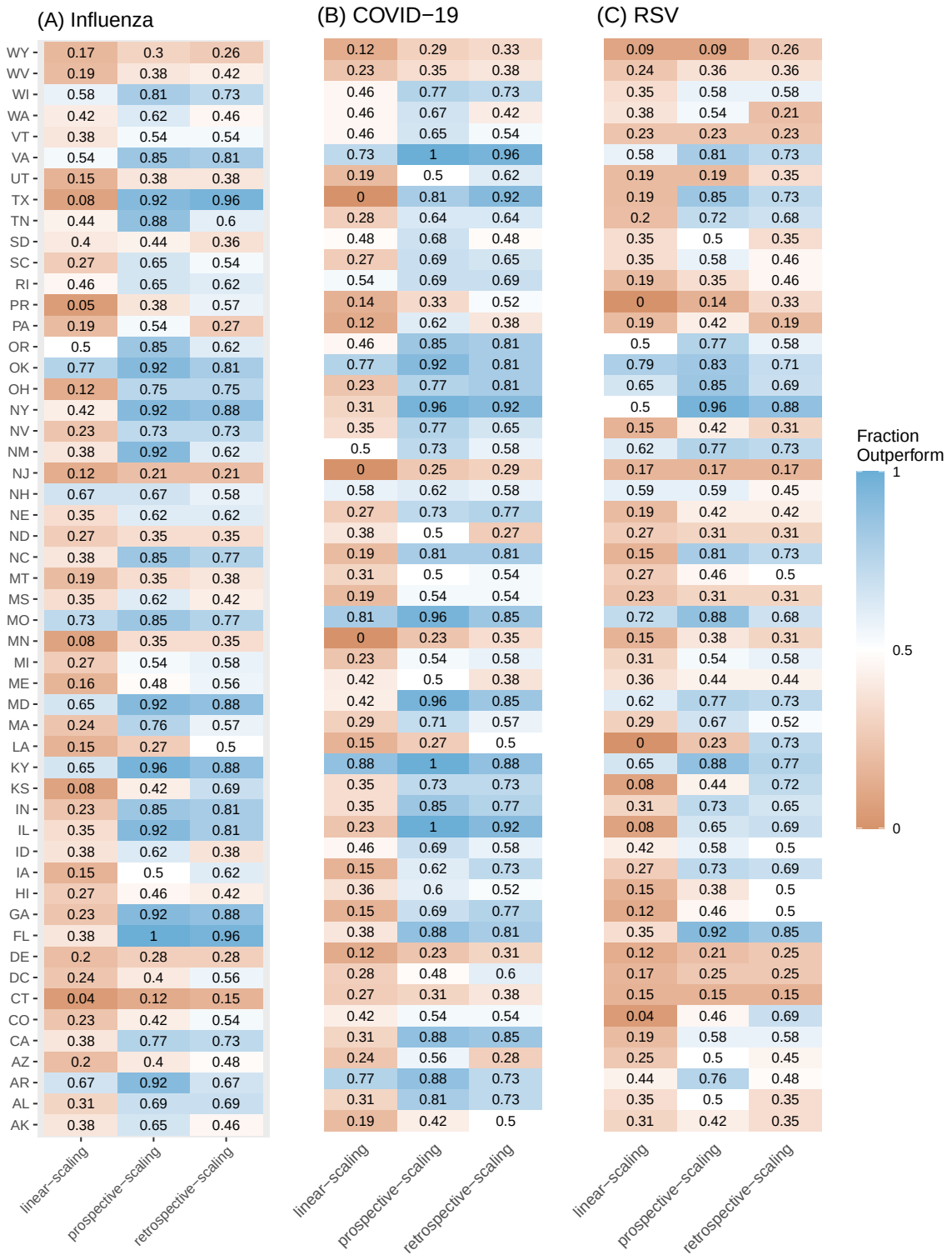


Fig. 5 Fraction of weeks for each location (y-axis) in which a given model—linear-scaling, prospective-scaling, or retrospective-scaling (x-axis)—had a lower error than the NULL model. Panels (A), (B), and (C) show results for influenza, COVID-19, and RSV, respectively.

Table 2 The number of locations, for each model and pathogen, in which the model outperformed the NULL model for more than half of the analysis period.

	Influenza	COVID-19	RSV
Linear-scaling	8	7	8
Prospective-scaling	32	37	24
Retrospective-scaling	33	38	23

5 Discussion

Beginning in the 2024–25 respiratory season, the NHSN dataset introduced a preliminary release each Wednesday containing weekly reported data. Historically, the last data point in a preliminary release is characterized by a lower reporting fraction and is subsequently updated in both the first and later finalized data releases. Accurately estimating the true value of this last data point is important for both public health officials and modelers.

In the absence of historical backfill statistics for this final data point, we developed and implemented a simple algorithm to prospectively estimate the extent of backfill during the 2024–25 respiratory season. In this retrospective analysis, we compared the performance of the prospective-scaling model to that of a NULL model and two alternative models: linear-scaling and retrospective-scaling. The prospective-scaling model can be viewed as a compromise between the NULL model, which assumes no backfill, and the linear-scaling model, which assumes that backfill scales linearly with the fraction of hospitals that have not yet reported. The retrospective-scaling model, by contrast, requires access to data for the complete season and is therefore not suitable for real-time use.

Our findings indicate that the intuitive linear-scaling assumption performs poorly when applied to the NHSN data. It consistently overestimates the extent of backfill, resulting in larger errors than the NULL model across the majority of location-week combinations for all three pathogens. Both the prospective-scaling and retrospective-scaling models outperformed the NULL model in most location-week combinations for influenza and COVID-19, but not for RSV, and both models demonstrated similar performance across all three pathogens.

Given that all three pathogens followed the same reporting structure during the study period, i.e. the same initial and final reporting fractions, we speculate that the lower performance of the models for RSV may be due to its lower overall hospitalization burden compared to influenza and COVID-19. This reduced signal may make it more difficult for scaling-based methods to provide accurate adjustments.

A core assumption of the prospective-scaling model is that backfill is non-negative. Although our analysis excluded the 11 locations-week instances in which the preliminary reporting fraction

decreased, the actual backfill was not always positive, which negatively impacted the prospective-scaling model's performance. For example, Arizona (AZ), Louisiana (LA), and Montana (MT) each had 10 weeks of negative influenza backfill (out of 26), and in these cases, the prospective-scaling model underperformed relative to the NULL model. Similarly, Minnesota (MN) had 8 weeks of negative backfill and showed similar trends (see Table 1 for a complete list of the number of weeks by location with negative backfill).

Because this was the first respiratory season with the current reporting cadence and requirements, and because most instances of negative backfill occurred early in the season, we speculate that this issue may be less prevalent in future seasons. This, in turn, suggests that the retrospective-scaling model, which is linearly optimized for the 2024-25 backfill pattern, may not generalize well to future seasons.

The nearly uniform overestimation of backfill by the linear-scaling model is interesting and non-intuitive. It may suggest that the catchment areas of facilities that are late to report are smaller than those of facilities that report earlier. This interpretation relies on the assumption that prevalence and age distribution are similar across all reporting facilities, an assumption that is unlikely to hold in practice.

A notable outlier in the reporting pattern is the state of Iowa (IA). Here, the extent of backfill was positively correlated with the initial reporting fraction, a counterintuitive result. We are unable to explain this anomaly without additional state-level information, such as individual facility-level reporting patterns and data.

Our simple scaling model has several limitations. First, it assumes that backfill is always positive, a condition that failed more often than expected during the 2024–25 respiratory season. Second, the model applies the scaling only to the last preliminary reported data point, whereas earlier weeks may also exhibit some degree of backfill. Although the final data point typically experiences the largest revisions, ignoring earlier weeks may overlook important corrections. Third, the model provides only a point estimate, without any upper or lower bounds or a full probabilistic distribution.

Despite these limitations, our analysis suggests that in the absence of negative backfill (which we expect to be less frequent in the next respiratory season), the prospective-scaling is likely to yield the lowest error in real-time influenza and COVID-19 applications, and that the NULL and linear scaling models can serve as reasonable lower and upper bounds, respectively. In contrast, for RSV, the NULL model, assuming no backfill, appears to provide the lowest backfill error.

6 Conclusions

This retrospective study evaluated a simple, operational nowcasting procedure for estimating the true value of the most recent hospital–admissions data point in the NHSN Wednesday preliminary release.

The procedure requires only current-season reporting–fraction information for each jurisdiction and is applicable to influenza, COVID-19, and RSV.

Throughout the 2024–2025 season, this prospective scaling approach outperformed both a NULL (no-scaling) model and a linear-scaling model for influenza and COVID-19. In contrast, for RSV the NULL model yielded the lowest error. The retrospective linear-regression model performed comparably to the prospective model, but it may not perform well in future seasons given that it was calibrated using only a single season.

Future work will extend the point-estimate prospective scaling into a probabilistic framework that captures uncertainty.

Supplementary information. An Appendix with seven figures is included at the end of this document (and not as a separate file.)

7 List of Abbreviations

CDC Centers for Disease Control and Prevention

CSTE Council of State and Territorial Epidemiologists

NHSN National Healthcare Safety Network

RSV Respiratory Syncytial Virus

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The Weekly Hospital Respiratory Data Metrics by Jurisdiction, National Healthcare Safety Network dataset analyzed in this manuscript is publicly available from: [https://data.cdc.gov/d/mpgq-jmmr] (preliminary-Wednesday release) and [https://data.cdc.gov/d/ua7e-t2fy] (official-Friday release). This analysis relies heavily on the state of these datasets as they were reported each week of the 2024-25 respiratory season. To our knowledge, these data are not publicly available. The authors have published their unofficial archive of these data on [https://github.com/predsci/nhsn-backfill-manuscript.git]. All code used to generate the figures presented in this manuscript is publicly available on GitHub at [https://github.com/predsci/nhsn-backfill-manuscript.git].

Competing interests

MBN, JT, and PR have wages paid by and financial interest in the company Predictive Science Inc. However, no current or past projects, relationships, or opportunities can be considered a competing interest to the subject of this manuscript.

Funding

This work was supported by Council of State and Territorial Epidemiologists (CSTE) under cooperative agreement number NU38OT000297. CSTE did not have a role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

MBN contributed to the study conception, data analysis and interpretation, and development of the analysis and plotting code, and drafted and revised the manuscript. JT collected the data, and contributed to the study conception, interpretation of data, and manuscript development. PR aided in the interpretation of the data and model fits and assisted in the development of the manuscript.

Acknowledgments

The authors thank Drs. Rebecca Borchering and Matthew Biggerstaff for insightful discussions that improved this work.

References

[1] Nixon K, Jindal S, Parker F, Marshall M, Reich NG, Ghobadi K, et al. Real-time COVID-19 forecasting: challenges and opportunities. The Lancet Digital Health. 2022;4(10):e699–e701. https://doi.org/{10.1016/S2589-7500(22)00167-4}.

[2] Lopez VK, Cramer EY, Pagano R, Drake JM, O'Dea EB, Adee M, et al. Challenges of COVID-19 Case Forecasting in the US, 2020–2021. PLOS Computational Biology. 2024;20(5):e1011200. https://doi.org/10.1371/journal.pcbi.1011200.

[3] Kim TH, Chinthaginjala R, Srinivasulu A, Tera SP, Rab SO. COVID-19 health data prediction: a critical evaluation of CNN-based approaches. Scientific Reports. 2025;15:9121. https://doi.org/10.1038/s41598-025-92464-0.

[4] U S Department of Health and Human Services, Centers for Disease Control and Prevention.: Weekly United States Hospitalization Metrics by Jurisdiction, National Healthcare Safety Network (NHSN). Accessed: 2026-01-14. https://healthdata.gov/CDC/

Weekly-United-States-Hospitalization-Metrics-by-Ju/kruc-9unj.

[5] Scobie HM. Correlations and Timeliness of COVID-19 Surveillance Data Sources and Indicators—United States, October 1, 2020–March 22, 2023. MMWR Morbidity and Mortality Weekly Report. 2023;72.

[6] Reich NG, Brooks LC, Fox SJ, Kandula S, McGowan CJ, Moore E, et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. Proceedings of the National Academy of Sciences. 2019;116(8):3146–3154. https://doi.org/10.1073/pnas.1812594116.

[7] Lutz CS, Huynh MP, Schroeder M, Anyatonwu S, Dahlgren FS, Danyluk G, et al. Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. BMC Public Health. 2019;19:1659. https://doi.org/10.1186/s12889-019-7966-8.

[8] Mathis SM, Webber AE, León TM, Murray EL, Sun M, White LA, et al. Evaluation of FluSight influenza forecasting in the 2021–22 and 2022–23 seasons with a new target laboratory-confirmed influenza hospitalizations. Nature communications. 2024;15(1):6289. https://doi.org/10.1038/s41467-024-50601-9.

[9] CDC FluSight.: Forecasts of Flu Hospital Admissions. https://www.cdc.gov/flu-forecasting/index.html. Accessed 31 July 2025.

[10] Ma L, Qiu Z, Van Mieghem P, Kitsak M. Reporting delays: A widely neglected impact factor in COVID-19 forecasts. PNAS Nexus. 2024;3(6):pgae204. https://doi.org/10.1093/pnasnexus/pgae204.

[11] Tang J, Rumack A, Wilder B, Rosenfeld R. Real-time forecasting of data revisions in epidemic surveillance streams. PLoS Computational Biology. 2025;21(11):e1013709. https://doi.org/10.1371/journal.pcbi.1013709.

[12] Cramer EY, Huang Y, Wang Y, Ray EL, Cornell M, Bracher J, et al. The United States COVID-19 forecast hub dataset. Scientific data. 2022;9(1):462.

[13] Centers for Disease Control and Prevention (CDC) & Infectious Disease Dynamics Group.: RSV Forecast Hub. Accessed: 2026-01-14; Collaborative probabilistic forecasts of weekly RSV hospitalizations. https://rsvforecasthub.org/.

[14] Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J, Hassanien AE. Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN computer science.

2020;1(4):197.

[15] Fox SJ, Lachmann M, Tec M, Pasco R, Woody S, Du Z, et al. Real-time pandemic surveillance using hospital admissions and mobility data. Proceedings of the National Academy of Sciences. 2022;119(7):e2111870119. https://doi.org/10.1073/pnas.2111870119. https://www.pnas.org/doi/pdf/10.1073/pnas.2111870119.

[16] Mahajan S, Heller MT, Xu Y, Chan P, et al. Evaluation of Hospitalization Surveillance Systems for COVID-19 Hospital Admissions in the United States. Journal of Public Health Management and Practice. 2021;27(5):522–529. https://doi.org/10.1097/PHH.0000000000001308.

[17] Rhoads DD, Plummer P, Jones M, et al. Rapid Deployment of SARS-CoV-2 Testing and Surveillance Using State Public Health Laboratories. Public Health Reports. 2021;136:71S–83S. https://doi.org/10.1177/00333549211032666S03.

[18] Centers for Disease Control and Prevention.: NHSN Hospital Respiratory Data (HRD) Dashboard. Accessed: 2026-01-14. https://www.cdc.gov/nhsn/psc/hospital-respiratory-dashboard.html.

[19] Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. American Journal of Epidemiology. 2013;178(9):1505–1512. https://doi.org/10.1093/aje/kwt133.

[20] McGough SF, Johansson MA, Lipsitch M, Menzies NA. Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking. PLOS Computational Biology. 2020;16(4):e1007735. https://doi.org/10.1371/journal.pcbi.1007735.

[21] Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs H, Bosse N, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. Wellcome Open Research. 2020;5:112. https://doi.org/10.12688/wellcomeopenres.16006.1.

[22] Bosse NI, Abbott S, EpiForecasts, Baguelin M, Barlow J, Biggerstaff M, et al. Real-time COVID-19 forecasts: model performance and assessment of uncertainty. Epidemics. 2020;33:100356. https://doi.org/10.1016/j.epidem.2020.100356.

[23] Hogan AB, Jewell CP, Sherrard-Smith E, Vesga JF, Watson OJ, Whittaker C, et al. Nowcasting and forecasting the spread of COVID-19 and healthcare resource use: a multistate model with reporting delay. BMC Medicine. 2021;19:83. https://doi.org/10.1186/s12916-021-01984-z.

[24] Greene SK, McGough SF, Culp GM, Graf LE, Lipsitch M, Menzies NA, et al. Nowcasting for Real-Time COVID-19 Tracking in New York City: An Evaluation Using Reportable Disease Data From Early in the Pandemic. JMIR Public Health and Surveillance. 2021;7(1):e25538. https://doi.org/10.2196/25538.

Appendix A Supporting Information

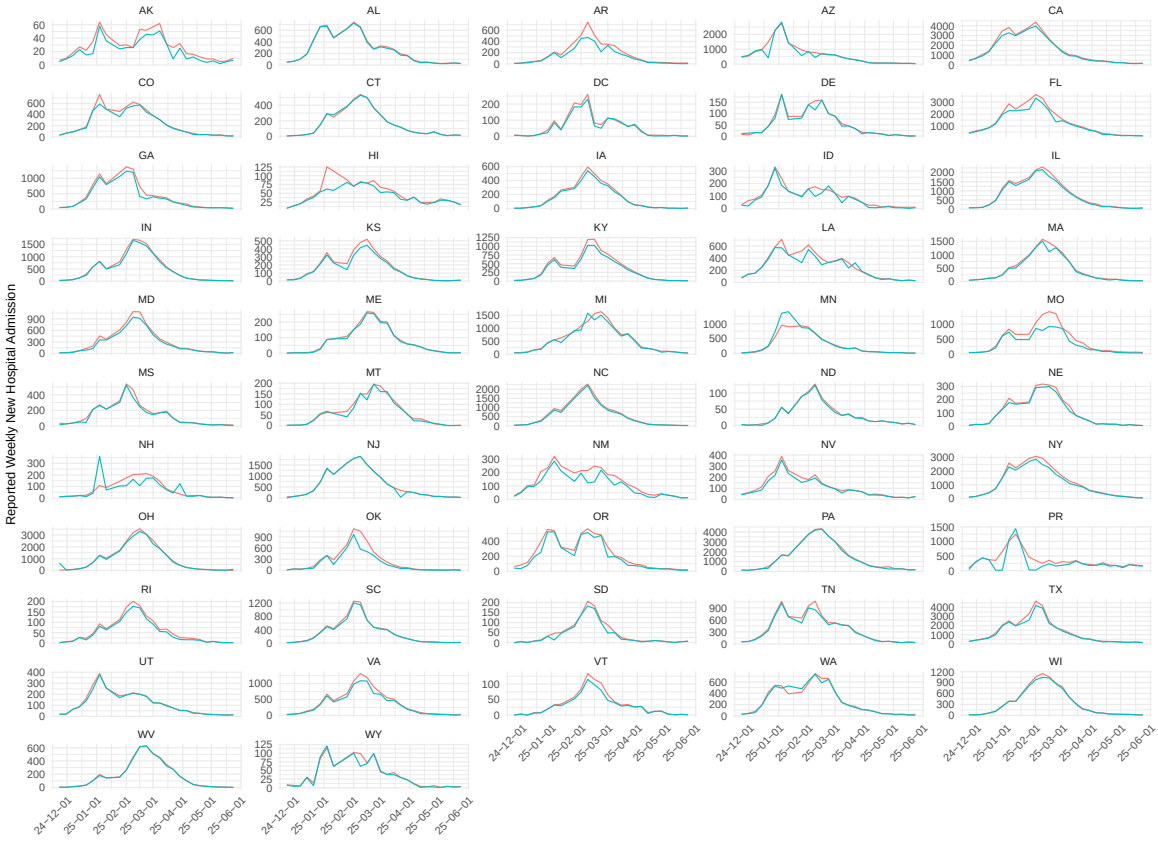


Fig. A1 Initial (green) and final (orange) weekly total new hospital admissions with confirmed influenza for all 52 jurisdictions and weeks (26) included in the study.

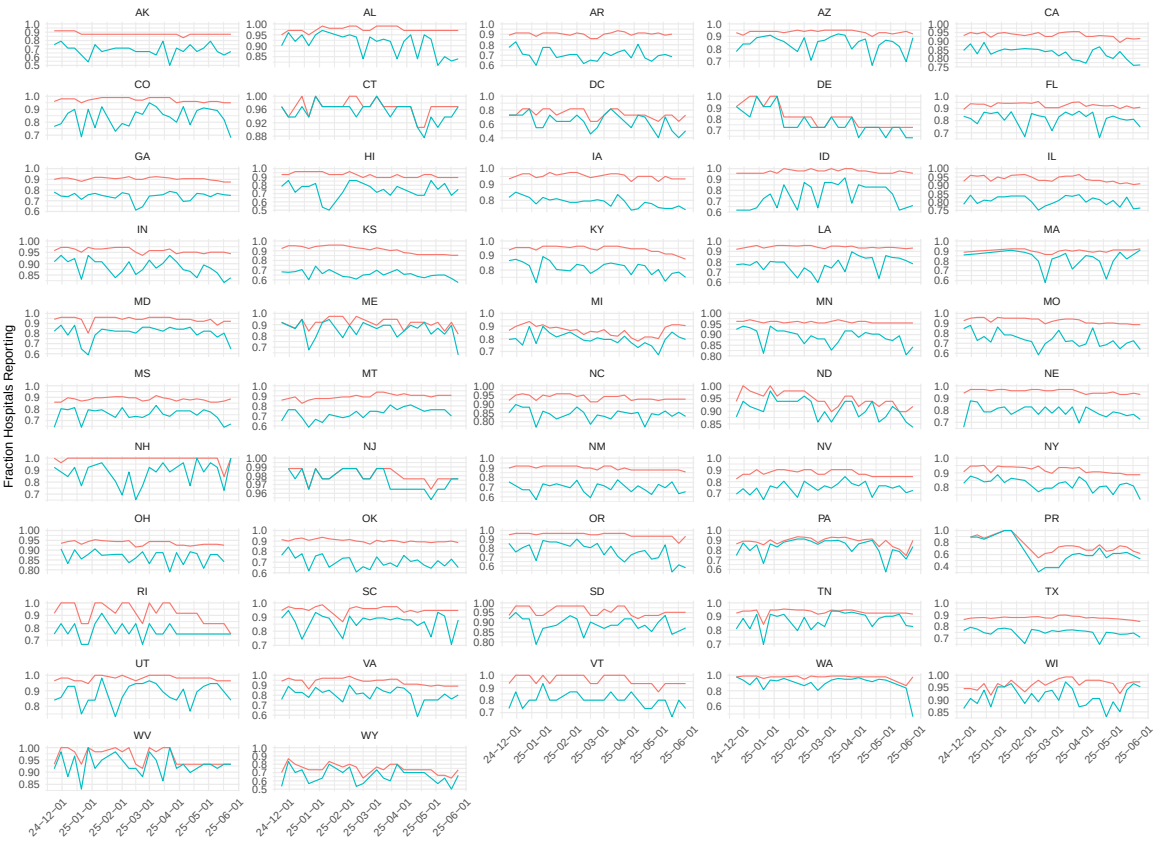


Fig. A2 Initial (green) and final (orange) fraction of reporting facilities for 52 jurisdictions and weeks (26) included in the study.

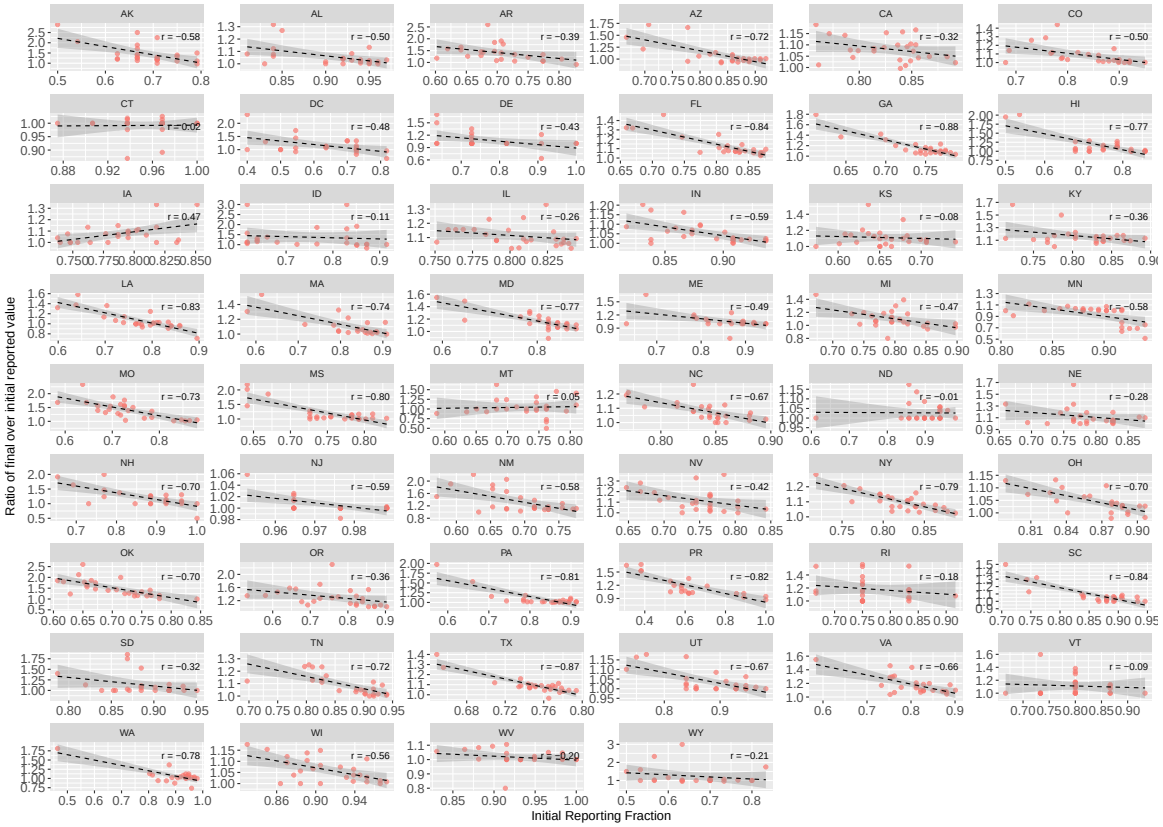


Fig. A3 Currently reported value divided by the initially reported value versus the initial reporting fraction for all 52 locations. The legend in each panel indicates the correlation between the two variables; the dashed line represents the linear regression fit, and the shaded area shows the 95% confidence interval.

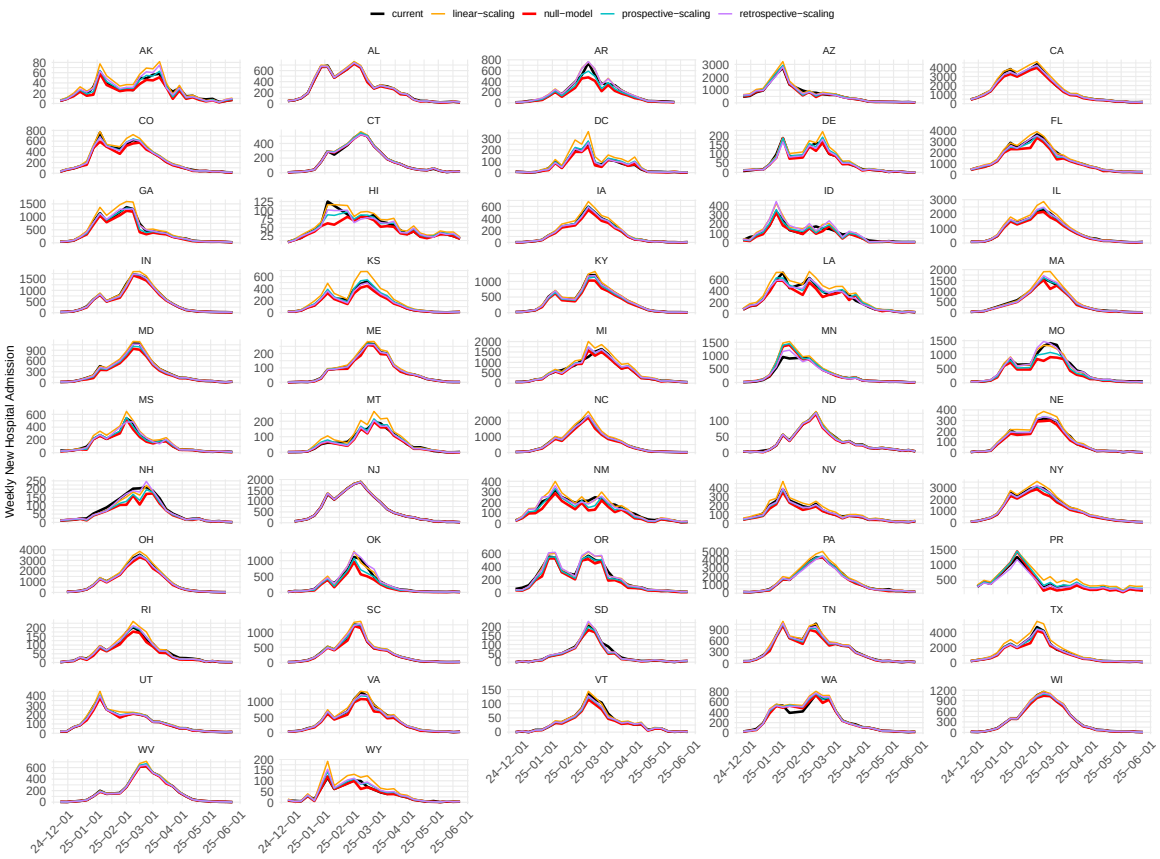


Fig. A4 Reported and models time-series for all 52 locations. Red line: current value, olive-green line:linear-scaling model, green line: NULL model (corresponding to the initial reporting of each value), blue line: prospective-scaling, magenta line: retrospective scaling.

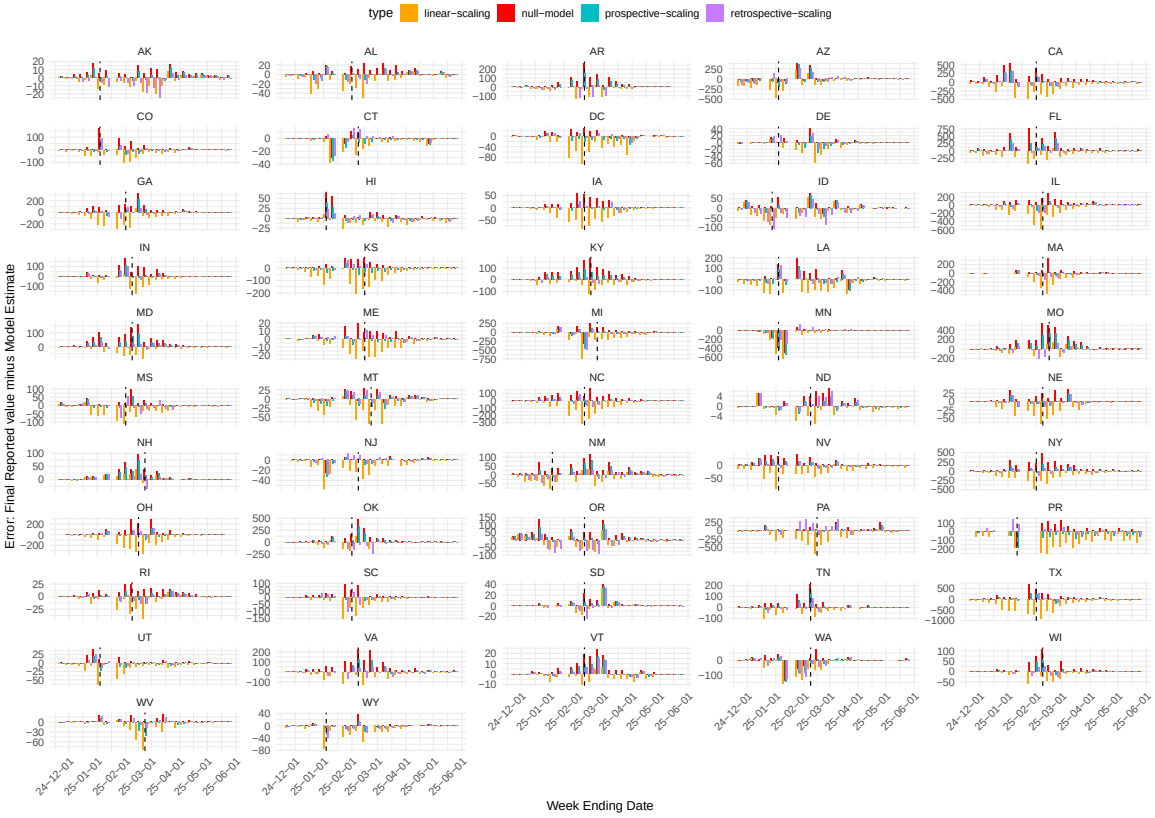


Fig. A5 The error of the NULL model (red), linear-scaling model (orange), prospective-scaling model (blue), and retrospective-scaling model (purple) for all 52 locations. The errors are calculated as the difference between the final reported value and the model estimate

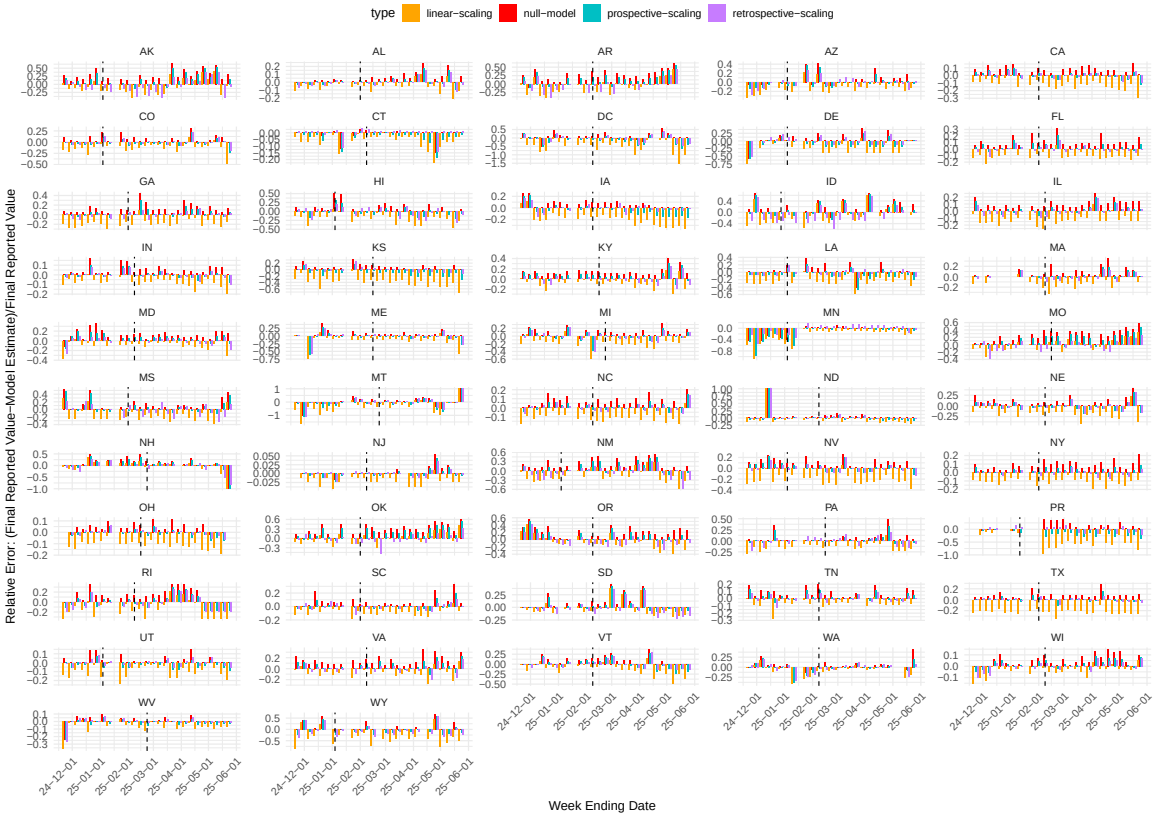


Fig. A6 The relative error of the NULL model (red), linear-scaling model (orange), prospective-scaling model (blue), and retrospective-scaling model (purple) for all 52 locations. The relative errors are calculated as the difference between the final reported value and the model estimate divided by the final reported value

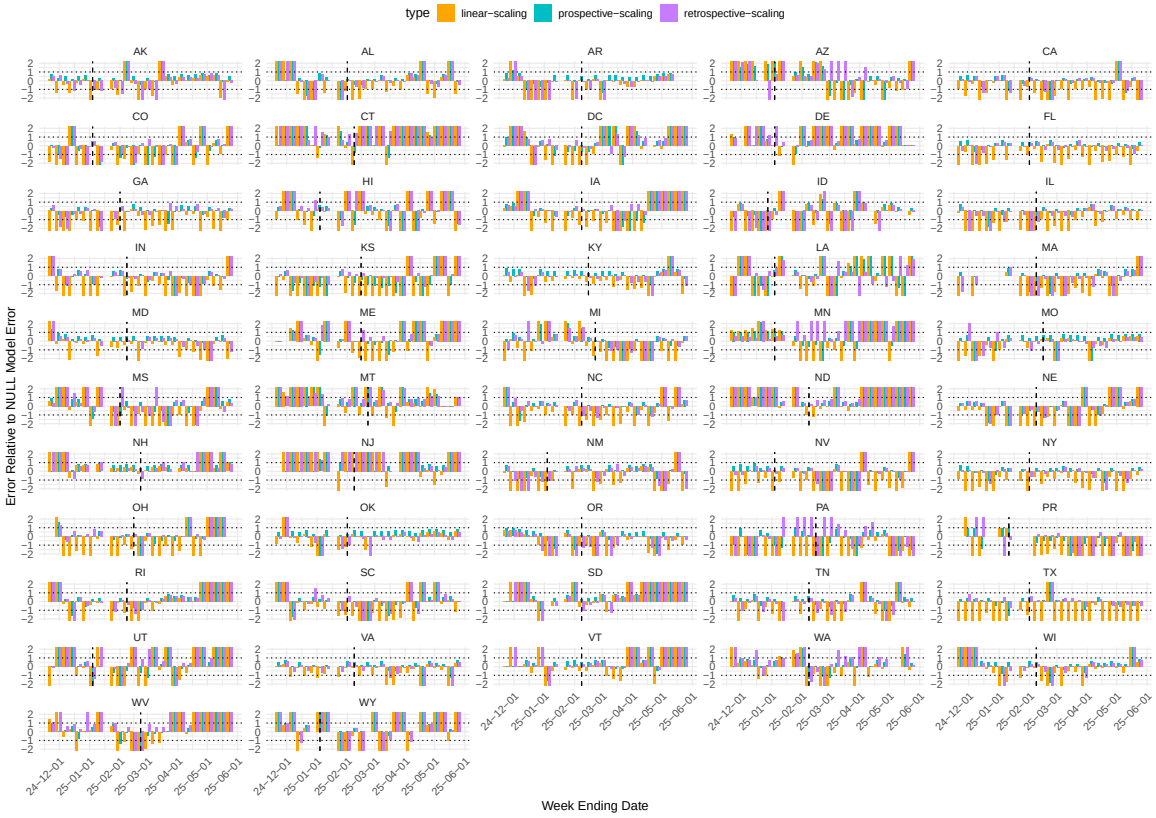


Fig. A7 The error relative to the NULL model error, linear-scaling model (orange), prospective-scaling model (blue), and retrospective-scaling model (purple) for all 52 locations. The error relative to the NULL model is computed by dividing the model's error by the error of the NULL model for the same location-week