FORECAST VS. REALITY: CDC FLUSIGHT PERFORMANCE IN THE 2024–2025 FLU SEASON

Andrea Thompson | INFO 610 | Fall 2025

INTRODUCTION

Background

Influenza and pneumonia are major causes of morbidity and mortality in the United States, ranking twelfth among leading causes of death between 2018 and 2023. Statistical forecasting plays a central role in preparing for each flu season—from vaccine strain selection to anticipating hospital burden. Because influenza transmission dynamics vary widely from year to year, accurate and well-calibrated models are essential.

FluSight Overview

Since 2013–2014, the CDC has run FluSight, a national forecasting initiative that synthesizes predictions from academic, governmental, and industry modeling groups into a weekly ensemble forecast of influenza hospitalizations nationally and in each US state and territory. These forecasts include 50% and 95% prediction intervals (PIs) intended to quantify model uncertainty. Evaluating how often observed hospitalizations fall inside these intervals provides insight into whether the model is well-calibrated or systematically over- or underpredicting.

Research Question

Did influenza hospitalizations during the 2024–2025 season fall within the CDC FluSight 50% and 95% prediction intervals at the expected rates?

Hypotheses

Let w_{50} = proportion of observations inside the 50% PI.
Let w_{95} = proportion inside the 95% PI.

50% PI
- H_0: $w_{50} = 0.5$
- H_a: $w_{50} > 0.5$

95% PI
- H_0: $w_{95} = 0.95$
- H_a: $w_{95} > 0.95$

DATA & METHODS

Data Acquisition

I downloaded weekly FluSight ensemble forecasts with one, two, and three week horizons and observed influenza hospitalization counts from the CDC Epidemic Prediction Initiative GitHub repository.

Processing

I merged forecast quantiles with observed outcomes for all US states/territories and the national level. I created two binary accuracy indicators:
- correct_50 = 1 if the observed value fell within the 50% PI
- correct_95 = 1 if the observed value fell within the 95% PI

Analytic Approach

- Calculated national and state-level coverage for the 50% and 95% prediction intervals.
- Standardized state accuracy using z-scores and grouped states into SD-based deviation categories to compare performance across regions.
- Evaluated point accuracy using a simple linear regression of observed hospitalizations on the median forecast.
- Produced time-series panels for four representative locations (National, California, Texas, Vermont) to illustrate differences in volume, seasonality, and variation.

Software

All analyses and visualizations were made in RStudio.

MODEL PERFORMANCE

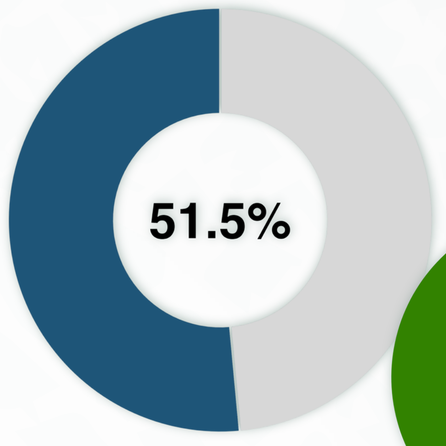
Linear Regression Summary

Statistic	Value	Interpretation
Slope (β_1)	**1.18*** **	Forecast underpredicts during high-volume weeks ($\beta > 1$)
Intercept (β_0)	**48.5**	Small baseline offset
Adjusted R^2	**0.901**	Explains ~90% of observed variation
Residual SD	**1013**	Typical magnitude of forecast error

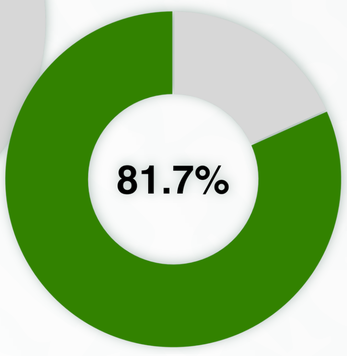
PROPORTION OF OBSERVATIONS WITHIN THE PREDICTION INTERVAL

50% Prediction Interval

51.5% of all observations fell within the 95% PI, slightly exceeding the nominal 50% expectation.



95% Prediction Interval

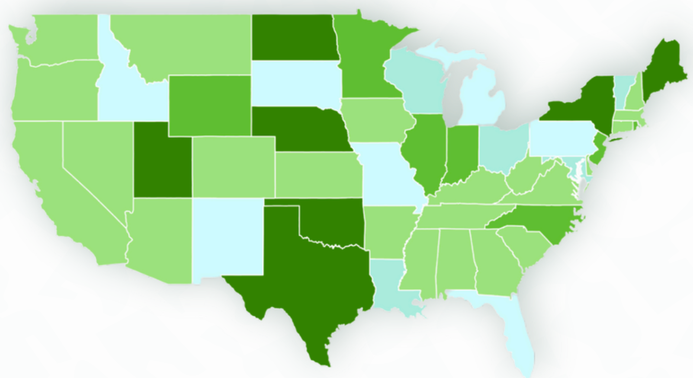


81.7% of all observations fell within the 95% PI, which is substantially below the nominal 95% expectation.

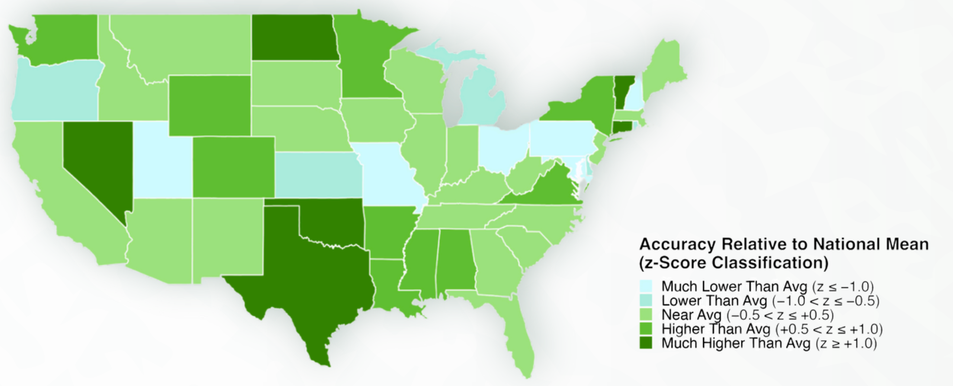
DEVIATION OF STATE-LEVEL FORECAST ACCURACY FROM THE NATIONAL MEAN

Z-Score Classification for 50% and 95% Prediction Interval Coverage

50% Prediction Interval



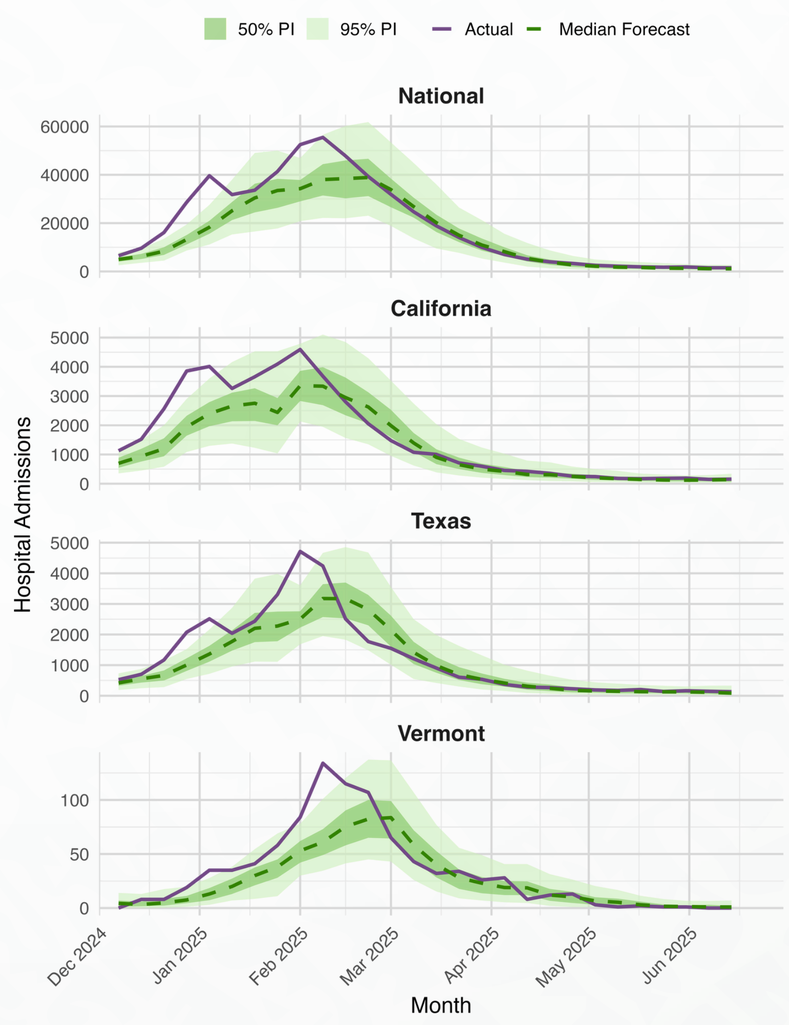
95% Prediction Interval



Forecast accuracy was uneven across states. 50% PI accuracy covered ~40–70% of observations, and 95% PI accuracy covered ~75–90%. Some Northern Plains and Northeast states achieved notably higher accuracy, while lower accuracy clustered in parts of the South and West. Applying z-score SD bands (≤ –1.0, –1.0 to –0.5, etc.) highlights these regional differences in a consistent, comparable framework.

FORECAST VS. ACTUAL HOSPITAL ADMISSIONS

50% and 95% Prediction Intervals for Selected Locations



I chose locations that represent diverse forecasting contexts: national-level performance, a large state (California), a highly seasonal state (Texas), and a small-population state (Vermont).

CONCLUSION & LIMITATIONS

The CDC FluSight ensemble performed well overall. The 50% prediction interval was almost perfectly calibrated, indicating the model generally captured typical week-to-week variation. In contrast, the 95% interval was too narrow, especially during higher-incidence weeks, pointing to an underestimation of uncertainty when activity rises quickly.

Point Accuracy

The median forecast explained roughly 90% of the variation in observed hospitalizations, which is a strong result. However, the model consistently underpredicted during rapid increases in flu activity.

State-Level Variation

Forecast accuracy varied across states. Differences may reflect population size, reporting practices, and regional flu dynamics. These structural patterns highlight where the model performs reliably and where additional calibration may be necessary.

Limitations

This analysis covers only the 2024–2025 season. Z-score comparisons assume relatively stable variance across states, which may not fully hold. The linear regression relies solely on the median forecast; other quantiles were not evaluated.

Scan for code, reproducible materials, and references

https://bit.ly/FluSightLinks