**Abstract**

Accurately estimating infectious disease hospitalizations at small geographic scales remains a significant challenge. Most forecasting efforts to date have focused on national or state-level predictions, leaving a gap in localized decision support. Yet, county- and region-specific estimates are critical for effective resource allocation and targeted public health interventions. In this study, we developed a flexible, data-driven framework to distribute statewide projections of respiratory-related hospitalizations—including those from influenza, COVID-19, and RSV—across counties and regions in South Carolina.

We integrated hospitalization data from the South Carolina Revenue and Fiscal Affairs (RFA) office with electronic health record (EHR) encounter data from two major health systems in South Carolina—MUSC and Prisma Health. Separate models were developed for COVID-19, influenza, and RSV. Counties with fewer than 500 total encounters in 2023 were excluded from the COVID-19 and influenza models, and counties with fewer than 100 encounters were excluded from the RSV model. To avoid overcounting, repeat admissions occurring within 14 days were removed. Weekly hospitalization counts were smoothed using a peak-preserving moving average algorithm. Gradient-boosted tree models (XGBoost) were trained on observed hospitalizations from 2020 to 2023, using predictors such as geographic area, epidemiological week, year, and total EHR encounters. Model hyperparameters were optimized using Bayesian Optimization. The final models were applied to distribute CDC-projected statewide hospitalizations for 2024–2025 to counties and regions based on predicted risk. Model performance was assessed using Percent Agreement (PA) between smoothed observed and distributed hospitalization counts.

Across all three diseases, region-level models consistently produced higher agreement with observed hospitalizations than county-level models. For COVID-19, county-level predictions achieved a median PA of 0.82 (IQR: 0.71–0.91), while region-level predictions performed better with a median PA of 0.93 (IQR: 0.90–0.95). The county-level median PA for influenza was 0.75 (IQR: 0.53–0.90), compared to 0.82 (IQR: 0.70–0.91) at the region level. For RSV, county-level predictions yielded a median PA of 0.78 (IQR: 0.57–1.00), while region-level predictions reached a median PA of 0.89 (IQR: 0.83–0.95).

This study demonstrates the feasibility of distributing statewide projections of respiratory-related hospitalizations to smaller geographic units using a flexible, data-driven framework. By leveraging EHR encounter data and machine learning models, we achieved a strong agreement with observed hospitalization patterns—particularly at the regional level. These methods can support more localized public health planning and resource allocation for current and future respiratory disease outbreaks.