Title: Small-Area Effective Reproductive Number Estimates and Machine Learning Forecasting of EHR Outcomes for Local Infectious Disease Preparedness.

Abstract:

Our research integrated advances in spatial effective reproductive number (R_t) estimation with local-level infectious disease forecasting to support granular surveillance across South Carolina. We developed a two-step framework for estimating R_t in small geographic areas where case data were sparse or noisy. We generated initial R_t estimates using 'R packages' such as EpiEstim, EpiFilter, and EpiNow2, and then refined them using a covariate-adjusted Bayesian INLA spatial model that incorporated sociodemographic predictors and spatial dependence. This approach yielded stable county- and ZIP-code–level R_t estimates and demonstrated strong agreement across diverse settings. We also developed a machine learning and probabilistic forecasting framework to predict R_t and daily infectious disease case counts, respectively. This study presents a flexible forecasting framework that integrates Bayesian estimation, spatial smoothing, and ensemble machine learning to improve the accuracy of disease transmission and case forecasts. In addition, we applied machine learning models to de-identified electronic health records to forecast weekly clinical outcomes, including encounters, emergency department visits, hospitalizations, and positive tests for COVID-19 and influenza. These models incorporated lagged predictors, temporal and seasonal features, holiday effects, and rate-of-change measures. The resulting forecasts closely matched observed trends and provided uncertainty estimates to support timely and data-driven public health decision-making.