Predictors of Test Positivity and Mortality During Four Waves of the Coronavirus Disease Pandemic in Orange County

STAT 295 Final Project

Damon Bayer

Winter 2022

1 Introduction

SARS-CoV-2 is a novel human coronavirus associated with high morbidity and mortality (Wu and McGoogan 2020). Community spread of the virus was first reported in December 2019 in Wuhan, China and spread to Orange County, CA, by January 2020 (Orange County Healthcare Agency 2020). By March 2022, Orange County had reported nearly 550,000 cumulative cases and nearly 7,000 deaths (Orange County Healthcare Agency, n.d.). Early media coverage in Orange County and nationally noted that the burden of COVID-19 was not evenly distributed throughout society, with the worst impacts seen in Black and Latino communities (Lin II, Carcamo, and Jiménez 2020). Rigorous assessment of factors related to test positivity, mortality, and seropositivity during the early stages of the pandemic in Orange County was performed in Parker et al. (2021). The authors found that cases were clustered among high income areas of the county very early on, but quickly shifted toward areas with lower socioeconomic status.

This analysis aims to expand the investigation presented in Parker et al. (2021) while accounting for spatial correlation and to further evaluate heterogeneity in test positivity and death rate at later stages in the epidemic.

2 Methods

2.1 Data

Case, test, and mortality data for this analysis come from line lists provided by the Orange County Health Care Agency. For the analysis, we aggregate the counts of cases, tests, and deaths by zip code and epidemic "wave." Figure 1 shows the four waves - "First," "Winter," "Delta," and "Omicron" overlaid on county-wide weekly test positivity and deaths. Each wave is characterized by a clear rise and subsequent decline in testing positivity. The windows defining the waves for deaths are shifted forward two weeks from the waves defining test positivity because deaths typically occur about two weeks after contracting the disease. The overall test positivity and death rate for each wave and zip code are presented in presented in 2. Modeling this data is the objective of our analysis. These plots reveal a pattern of higher test positivity and death rate in the north central area of the county, which is comprised of Santa Ana and Anaheim, in the first two pandemic waves. During the Delta and Omicron waves, the disease burden appears to be more evenly distributed.

Zip-code-level demographic data comes from the 2019 American Community Survey as accessed through the `tidycensus` R package (Walker and Herman 2022). This demographic data is displayed in Figure 3. Broadly, these plots show that most demographics vary in a radial pattern based on proximity to the cities of Santa Ana and Anaheim in the north central area of the county. Zip codes in these cities tend to be less educated, less insured, more Hispanic and Latino, lower income, and higher density than communities outside of this area. These patterns closely resemble the patterns in test positivity and death rate observed in the first two pandemic waves.
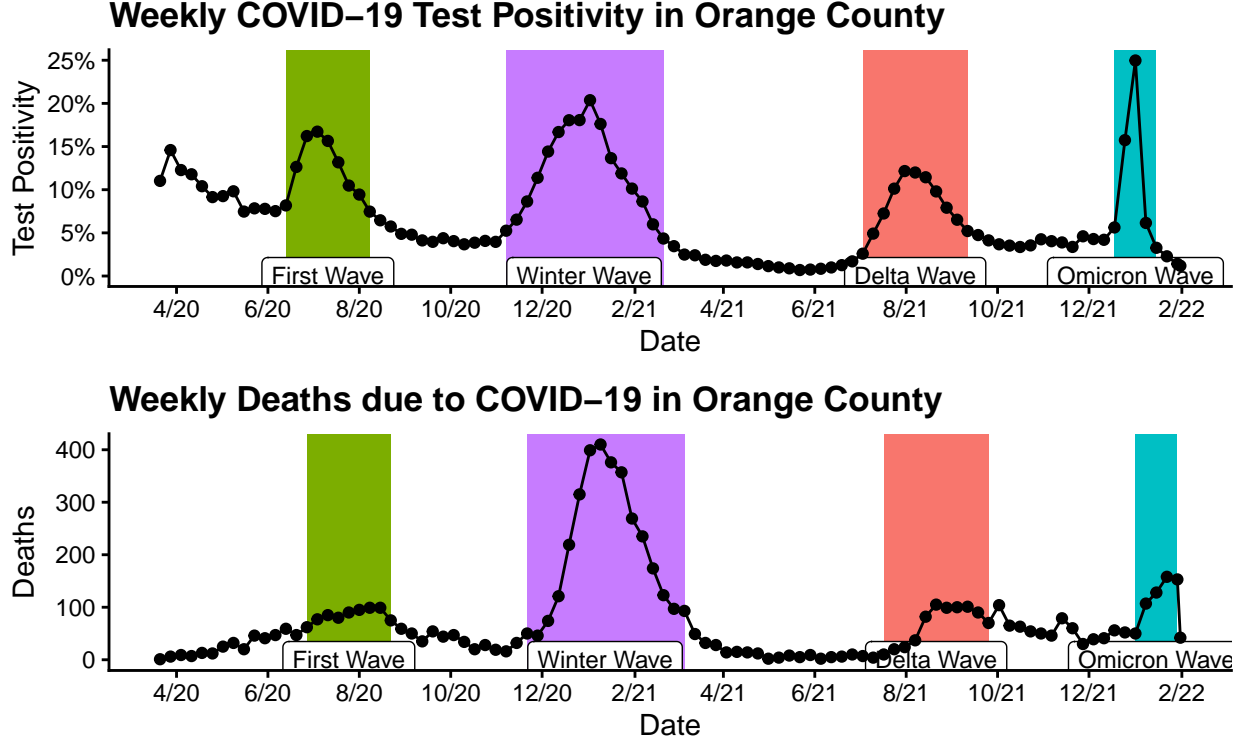
Figure 1: Weekly Test Positivity and Deaths in Orange County throught four epidemic waves - "First," "Winter," "Delta," and "Omicron." Death waves are offset from Test Positivity waves by two weeks.

2.2 Statistical Models

We take inspiration for our models from those proposed in Parker et al. (2021), using the same predictor and response variables, but with some important modifications:
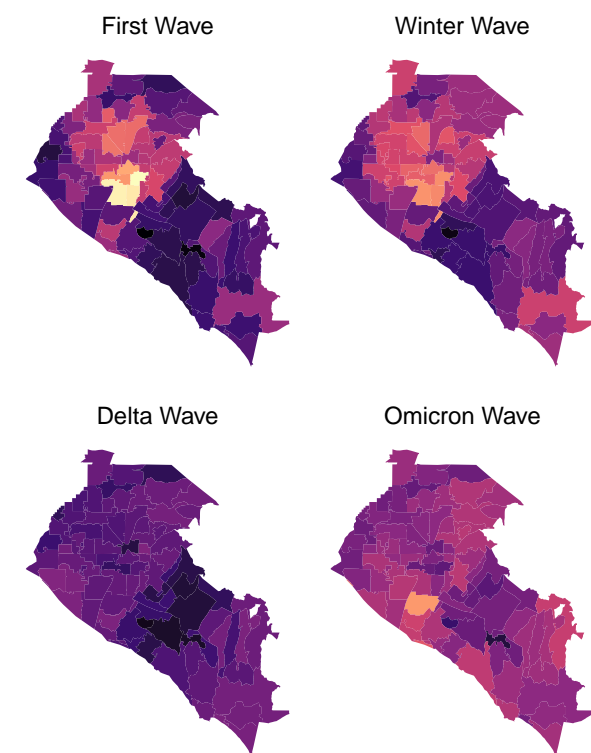
- Rather than using individual level data when available, we aggregate all data at the zip code level to create areal data.
- Rather than binning continuous independent variables, we treat them as continuous.
- Rather than modeling time continuously, we discretize into the First, Winter, Delta, and Omicron waves.
- We model spatial random effects with an improper conditional autoregressive model.

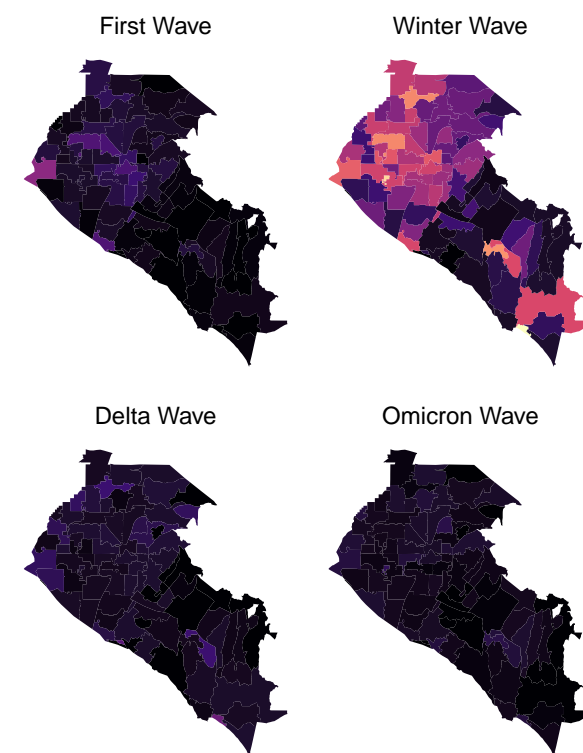The hierarchical model for testing positivity is:

$$\text{cases}_{i,t} \mid \pi_{i,t} \sim \text{Binomial}(\pi_{i,t} \, \text{tests}_{i,t})$$

$$\log\left(\frac{\pi_{i,t}}{1 - \pi_{i,t}}\right) = \mathbf{X}_i^T \boldsymbol{\beta}_t + \boldsymbol{\eta}_i$$

$$\eta_i \mid \{\eta_j, j \neq i\} \sim \text{N}\left(\frac{\sum_{j \neq i} w_{ij} \cdot \eta_j}{w_{i+}}, \frac{\tau^2}{w_{i+}}\right)$$

The hierarchical model for deaths is:

Test Positivity

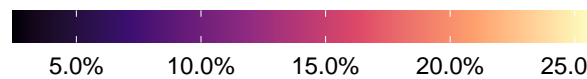## First Wave

Winter Wave

Death Rate per 100,000

First Wave

Winter Wave

Delta Wave

Omicron Wave

Delta Wave

Omicron Wave

Test Positivity

5.0%   10.0%   15.0%   20.0%   25.0%

Death Rate per 100,000

0   100   200

Figure 2: Test positivity rate and death rate for four pandemic waves in Orange County, CA

Age 65+

Age 85+

Have Bach. Degree

Have Health Insur.

%

20% 40% 60% 80%

%

0.0% 5.0% 10.0% 15.0% 20.0%

%

20% 40% 60% 80%

%

85.0% 90.0% 95.0%

Hispanic or Latino

Med. House Income

Pop. Density

%

20% 40% 60% 80%

$

5.0 10.0 15.0 20.0

People/sq. mi

0.40 0.80 1.20 1.60
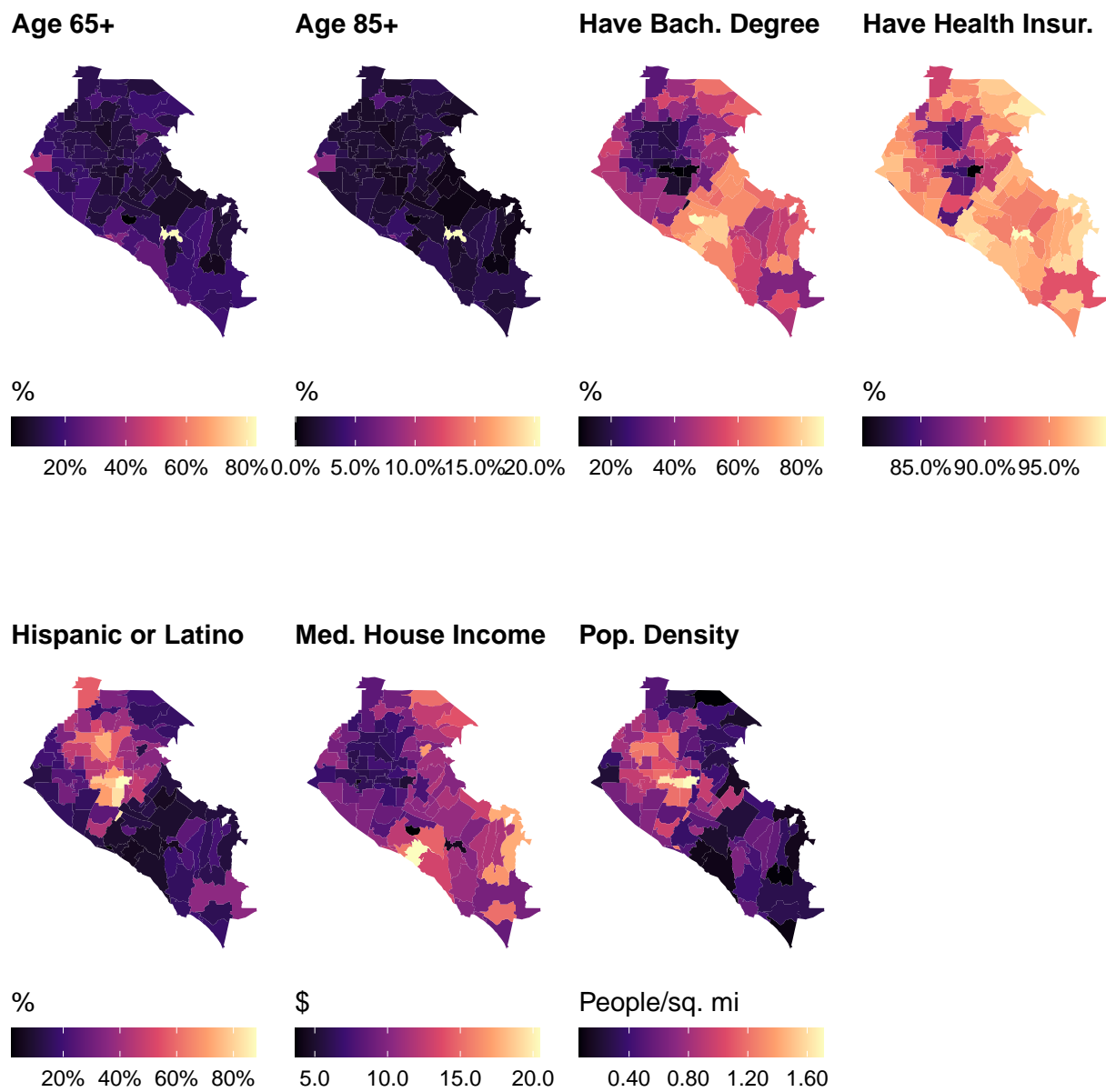
Figure 3: Demographic data for Orange County, CA.

$$\text{deaths}_{i,t} \mid \theta_{i,t} \sim \text{Poisson}(\theta_{i,t} \times \text{population}_i)$$

$$\log(\theta_i) = \mathbf{X}_i^T \boldsymbol{\beta}_t + \boldsymbol{\eta}_i$$

$$\eta_i \mid \{\eta_j, j \neq i\} \sim \text{N}\left(\frac{\sum_{j \neq i} w_{ij} \cdot \eta_j}{w_{i+}}, \frac{\tau^2}{w_{i+}}\right)$$

In each model, i indexes a zipcode and t indexes a pandemic wave. The linear portion of the models consists of $\mathbf{X}_i^T$ – the vector of a constant, percent older than 65, percent older than 85, percent with a bachelor's degree, percent with health insurance, percent Hispanic or Latino, median household income, and population density for zip code i, $\boldsymbol{\beta}_t$ – a vector of coefficients for these terms in wave t, and $\boldsymbol{\eta}_i$ – the spatial random effect for zip code i. The spatial random effects are further described by w_{ij} – a binary variable indicating if zip codes i and j are neighbors, w_{i+} – the total number of neighbors of zip code i, and τ^2 – a variance parameter.

In the software implementation, the model is fit with one design matrix $\mathbf{X}$, which includes interaction effects allowing allow us to estimate each wave's coefficients separately. This matrix is centered and scaled for computational convenience. The neighbor weight matrix was constructed using the `poly2nb` function from the `spdep` package using geometry data collected provided in `tidycensus` (Bivand, Pebesma, and Gomez-Rubio 2013).

We conducted Bayesian inference for this model with Stan via the `brms` package using default priors: improper priors for the coefficients, a scaled Student's t distribution for for the Intercept, and a half scaled Student's t distribution for τ (Bürkner 2017). We collected 1000 posterior samples each from four chains after 1000 warm up samples. Convergence and mixing were assessed using potential scale reduction factors, effective posterior sample sizes, and traceplots of model parameters.

3 Results

Inspection of our 4000 posterior samples did not indicate any issues with our model fit. All $\hat{R}$'s are less than 1.02. Each parameter's effective sample size was in the high hundreds or low thousands. Additionally, there was no noticeable spatial correlation in either of the models' residuals.

We present the posterior distributions of the regression coefficient in Figure 4. With respect to the test positivity model, we observe that the odds ratios associated with one unit increases in median household income, proportion age over 65, proportion of age over 85, and education are relatively consistent over the course of all four pandemic waves. In contrast, there is a noticeable shift in the odds ratios associated with population density, health insurance, and proportion of Hispanic or Latino residents, with these odds ratios being mostly greater than one in the first two waves, and less than one in the second two waves.

With respect to the death rate model, we observe consistency in the rate ratios associated with one unit increases in population density, household income, education, and ethnicity. In this model, the rate ratios for both age variables shift over time. Larger proportions of population over age 85 were associated with higher mortality early on, but this relationship weakened after the winter wave. We observe the opposite trend for the proportions of population over age 65.

Comparing the two models, we see that the factors associated with contracting the disease are not always the same as those related to death from the disease. In particular, we note that areas with large populations over 85 were not more likely to test positive, but were much more likely to die, especially in the first two pandemic waves We also observe that areas with larger Hispanic and Latino populations were much more likely to test positive, but not any more likely to die.

4 Discussion

We have presented two models which aim to investigate factors associated with COVID-19 test positivity and deaths due to COVID-19 in four waves of the pandemic, while accounting for spatial correlation.
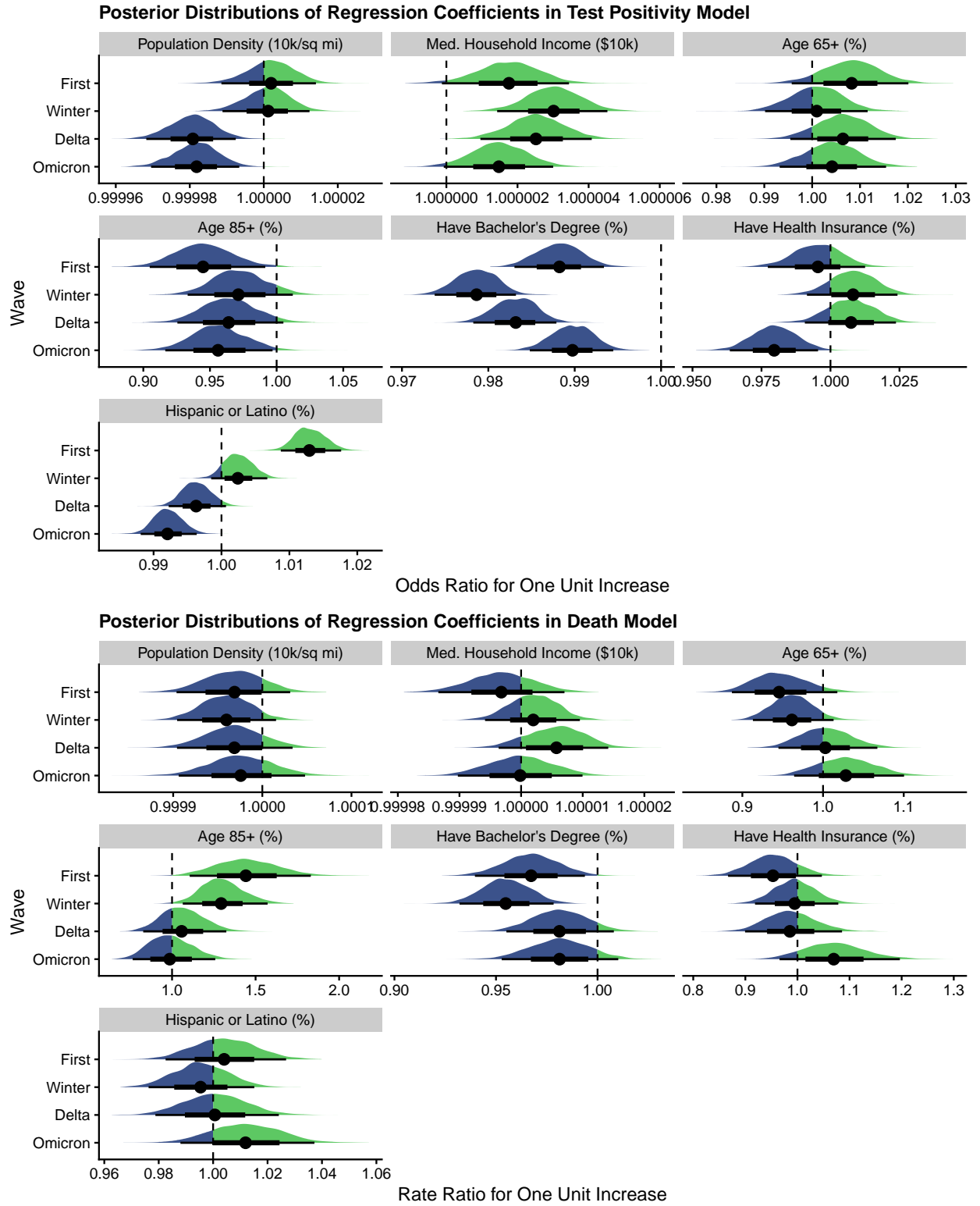
Figure 4: Posterior distributions of transformed regression coefficients for the testing positivity and death rate models.

Here, we offer some plausible explanations for the ratios which appear to change over time, based on our modeling results.

The changing direction for ratios associated with population density and proportion of Hispanic or Latino residence can likely be explained by the effects of natural immunity and increased transmissibility of the later coronavirus variants. In the first two waves, high density, heavily Hispanic and Latino areas were disproportionately affected, leading these communities to develop high levels of natural immunity to the virus, making them less susceptible to the Delta and Omicron variants. It is also possible that these areas benefited from target vaccination campaigns, which were enacted because they suffered so much prior to widespread vaccine availability.

The changing direction of the rate ratios associated with age structure in the deaths model can likely be explained by patterns in vaccination. The elderly benefit the most from vaccination because they are at such great risk of death due to COVID-19. Thus, they were highly motivated to get vaccinated and prioritized in vaccine distribution. After vaccinations became widely available after the winter wave, the vaccinated elderly were no longer at increased risk of death. A more pessimistic interpretation of this change is that the first two waves succeeded in killing so many vulnerable elders that there were many fewer left at risk in the latter waves. Because vaccine uptake was comparatively slower and overall less prominent in the 65+ age group, their risk of death increased, relative to the baseline rate, once all the very elderly were vaccinated.

For both models, we struggle to hypothesize a likely explanation for the changing ratios associated with health insurance.

While this model exercise has lead to valuable insights, there is room for improvement. The most obvious candidate for this would be a more sophisticated treatment of the temporal component of the data. Rather than modeling test positivity and deaths at four discretely binned times, it may be better to treat time as continuous. Additionally, it may be useful to allow the spatial random affects to vary over time.

Code to reproduce this analysis and paper is available at (https://github.com/damonbayer/stat_295_spatial_final.

References

Bivand, Roger S., Edzer Pebesma, and Virgilio Gomez-Rubio. 2013. *Applied Spatial Data Analysis with R, Second Edition.* Springer, NY. https://asdar-book.org/.

Bürkner, Paul-Christian. 2017. "brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80 (1): 1–28. https://doi.org/10.18637/jss.v080.i01.

Lin II, Rong-Gong, Cindy Carcamo, and Soudi Jiménez. 2020. "Coronavirus Takes Wrenching Toll on Orange County Latinos Who Have No Choice but to Work." *Los Angeles Times*, September. https://www.latimes.com/california/story/2020-09-26/coronavirus-positive-test-rates-much-higher-in-two-of-orange-countys-heavily-latino-cities.

Orange County Healthcare Agency. 2020. "OC Health Care Agency Confirms First Case of Novel Coronavirus in Orange County, California." https://mailchi.mp/ochca/novelcoronavirus.

———. n.d. "COVID-19 Case Counts and Testing Figures." https://occovid19.ochealthinfo.com/coronavirus-in-oc.

Parker, Daniel M, Tim Bruckner, Verónica M Vieira, Catalina Medina, Vladimir N Minin, Philip L Felgner, Alissa Dratch, Matthew Zahn, Scott M Bartell, and Bernadette Boden-Albala. 2021. "Predictors of Test Positivity, Mortality, and Seropositivity During the Early Coronavirus Disease Epidemic, Orange County, California, USA." *Emerging Infectious Diseases* 27 (10): 2604–18.

Walker, Kyle, and Matt Herman. 2022. *Tidycensus: Load US Census Boundary and Attribute Data as 'Tidyverse' and f'-Ready Data Frames.* https://CRAN.R-project.org/package=tidycensus.

Wu, Zunyou, and Jennifer M McGoogan. 2020. "Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases from the Chinese Center for Disease Control and Prevention." *JAMA* 323 (13): 1239–42.