STATS 295 Final Project - Anthony Hu and Spencer Matthews - December 2021

Spatio-temporal impact of self-financed rotavirus vaccination on rotavirus and acute gastroenteritis hospitalisations in the Valencia region, Spain - Lopez-Lacort et al. [1]

Background. The Rotavirus (RV) is a double-stranded RNA virus that causes gastroenteritis, the common stomach flu. The virus targets the small intestine lining and is primarily spread via the orofecal route through contaminated surfaces, hands, and objects. Before the Rotavirus vaccination program was initiated by the U.S. in 2006, acute gastroenteritis (AGE) was responsible for 60,000 hospitalizations and 37 deaths per year, most of which were children [2].

The World Health Organization (WHO) recommends two types of RV vaccinations for infants between 6 and 32 weeks of age. In European countries where RV vaccinations are compulsory, gastroenteritis related outpatient visits have fallen by 60-90% [3]. This large effect suggests that vaccination initiatives are practical and should be implemented in countries where distribution procedures are mature and streamlined.

This paper examines Spain, where RV vaccinations are unfunded by the National Health System. Though the country as a whole lags behind other European countries in vaccination rates, the Valencia region of Spain boasts medium vaccine coverage between 40-50% as defined by WHO recommendations. Given the heterogeneous distribution of infectious diseases, gastroenteritis related hospitalizations can be difficult to predict. Environmental and population characteristics can greatly alter the circulation of RV strains amongst susceptibles. Therefore, it is important to evaluate the ability of RV vaccinations to combat gastroenteritis for those most at risk.

Current estimation methods measure the impact of the RV vaccine by comparing gastroenteritis hospitalization trends in pre and post-authorization periods. This introduces heavy bias in final estimate totals [1]. The improvement introduced in this paper is to make use of spatio-temporal methods to reduce confounding errors that are present in many approximations of vaccine efficacy rates.

Problem and Mathematical/Statistical Formulation. This paper attempts to characterize vaccine efficacy in the Valencia region of Spain by predicting the number of hospitalizations that were avoided due to the RV vaccines. They do this by creating a spatio-temporal model of RV hospitalizations in each observation unit (where the units are broken down by vaccination status, age group, sex, biennial period, health department, and health district). The number of hospitalizations in each unit is modeled as $y_i \sim Bin(\theta_i, N_i)$ where i ranges over all 15,718 observational units. Each θ_i was modeled as shown in Figure 1, and each N was the population size (number of children) in each of the observational units.

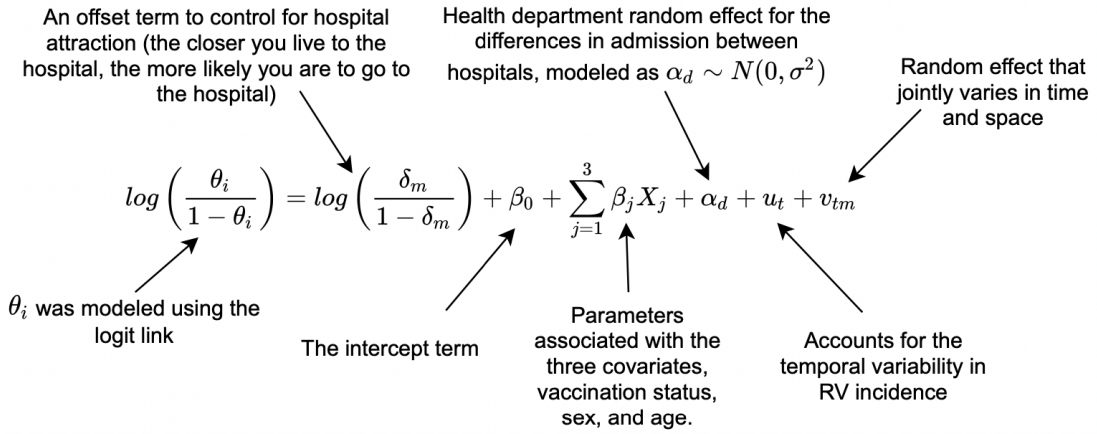

Figure 1: The model implemented to estimate the probability of hospitalization in each of the observation units, with a short description of each term.

They then used this model to predict the number of hospitalizations for each observational unit. To compute vaccine efficacy, the same predictions are computed but with the vaccination status set to "unvaccinated" for the entire population. The resulting predictions are then compared for each observational

unit, and subtracting the first model prediction from the model prediction where all children are considered unvaccinated gives and estimate of how many hospitalizations were averted due to the RV vaccine in this region of Spain over the given time period.

Three parameters account for spatial and temporal variability: u_t, δ_m, and v_{tm}. The biennial effect u_t considers correlation between adjacent time periods by a first order random walk modeled with an intrinsic conditional autoregressive (ICAR) prior distribution. The prior computes a gaussian risk factor based on the number of adjacent mappings in a connected graph, which in this case, represents adjacent time periods $(t-1)$ and $(t+1)$. In ICAR models, each connection has a weight which signal the random walk element, and in this model all the weights were set to one [4].

The random effect parameter $v_{t,m}$ follows a spatial-temporal autoregressive model. For the first biennial term, $v_{1,m}$ is formulated as

$$v_{1,m} = \frac{1}{\sqrt{1-\rho^2}} \cdot W_{1,m} \tag{1}$$

And each subsequent period (note that there are six total periods)

$$v_{t,m} = \rho v_{t-1,m} + W_{t,m}, \quad t = 2,3,4,5,6. \tag{2}$$

In Equations 1 and 2, ρ controls the temporal dependence and is assumed to have a uniform prior between -1 and 1. $W_{t,m}$ follows a spatial Besag, York and Mollie model where each observational unit represents a risk value formulated through a conditional Poisson model. The Poisson models takes in an expected count and a spatially-dependent risk factor from a Gaussian process [5]. Non-informative priors were considered for the β_j terms, and uniform priors between 0 and 5 were considered for all random effect standard deviations in the model.

The parameter δ_m measures the hospitalization rate for any cause. It follows a spatial Besag, York and Mollie model. If no other predictor in the model has an effect on θ_i, then the δ_m would equal θ_i. The δ_m's are estimated in a separate modelling process prior to the final model being fit, where the sampling process uses the form shown in Equation 3.

$$P(\delta_m|y_m) \propto P(y_m|\delta_m)P(\delta_m) \qquad \text{where } y_m \text{ is the hospital admissions in each region} \tag{3}$$

For this sampling step, 40,000 iterations were used, with half being discarded as burn-in. The mean of the resulting distribution was taken to be δ_m for the given region.

The overall model was fit using WinBUGS software to perform an MCMC analysis. This software was referenced by its corresponding R implementation R2WinBUGS in the code provided by the authors. The simulation was ran for 10,000 iterations with 2,000 being discarded as burn-in. Furthermore, only one out of every ten samples were saved. MCMC convergence was visually verified by examination of the posteriors.

Findings. In total, the study observed 17,482 RV-related hospitalizations. They considered 721,471 children, of whom 189,247 were vaccinated against RV. Vaccinated children accounted for 2,648 of the hospitalizations. In terms of population and setting demographics, two-year-old children were less likely to be hospitalized compared to infants less than one year old. Furthermore, girls were 19% less likely to be sent for RV-related hospitalizations than boys. Risk for acute gastroenteritis (AGE) trended downwards during the study while RV hospitalizations stabilized after 2010. Although vaccination rates differed greatly between health districts, overall vaccination coverage increased by 49% during the course of the study. With 189,247 vaccinated children, the fitted model estimated that 1,142 RV and 1,866 AGE hospitalizations were averted. This represents reductions of 19.9% and 10.2% of RV and AGE hospitalizations respectively. Assuming 100% vaccination coverage, RV hospitalizations would be lowered by 85.8% and AGE hospitalizations reduced by 46.9% compared to a completely unvaccinated population group.

Other Methods. Many spatio-temporal frameworks begin with defining a deterministic ODE model. One common approach is to fit time a series SIR (TSIR) model [6] to estimate a birth rate for new infectious individuals, β, and a death (recovery) rate, μ. Assuming independence from infectious individuals, this process is realized to be a negative binomial distribution. The TSIR model assumes that observed S_t and I_t are reasonable estimates. Given these observations, TSIR can compute a pure birth process by the conditional probability as shown in Equation 4.

$$I_t \sim NB(m_t, I_t), \qquad \text{where } m_t = \frac{\beta S_{t-1} I_{t-1}}{N} \tag{4}$$

From this process we can estimate a temporal-dependent infectious rate without introducing autoregressive ICAR priors. However, using TSIR with biennial time periods in problematic as time periods are often larger than our biennial term. For datasets with more granular time periods, TSIR can effectively estimate newly infected individuals with this simple process. Additionally, TSIR models can also be extended to account from spatial variability. In Equation 4, m_t can be extrapolated to m_{ti}, which accounts for an extra geographical location term i. As an example, the TSIR Gravity Model [7] introduces a penalization term in m_{ti} to register neighboring-area infectious individuals and corresponding movements between the regions of interest.

Numerical Reproductions and Application of Methods to COVID-19 Data. Before applying the methodology used in this paper to new data, we first reproduced the authors' work independently. Due to system requirements, we had to translate the provided WinBUGS code from the paper into the STAN language [8], which we accessed through the rstan package. We ran both the simulation to obtain δ_m as well as the simulation for the RV hospitalizations model, and obtained similar results to those reported in the paper. The values we obtained for δ_m were fairly close to those reported in the paper, with most of our estimated within 5% of the values saved in the paper repository. However, when we computed the number of averted RV hospitalizations caused by the vaccine, our estimate was a little bit different, giving a total of around 1,500 averted hospitalizations compared to the paper's reported value of 1,142. This discrepancy is likely caused by differences in final estimated parameters, since the variances associated with the parameters are quite large (especially in the β_j's), as shown by an example of β_3 in Figure 2. Due to the random nature of MCMC, differences in our final estimates could lead to the results we obtained.

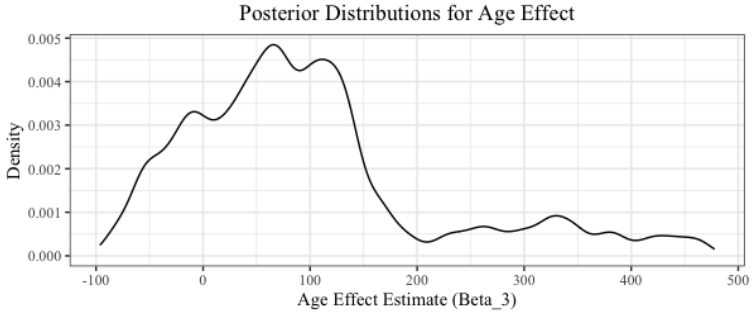

Figure 2: Posterior distribution for the Age Parameter, β_3 in the model for θ_i. This parameter seems quite dispersed and may have led to the difference between our estimated number of averted hospitalizations and the one reported in the paper.

One of our main concerns with the paper is that the model seems to contain too many parameters dealing with the spatio-temporal effect. In order to assess this in a way that was not possible in the published paper, we created posterior distributions for those parameters. The results for the time effect, u_t, can be seen in Figure 3. The posterior distributions in this plot do look fairly nice, except that they all seem to be very

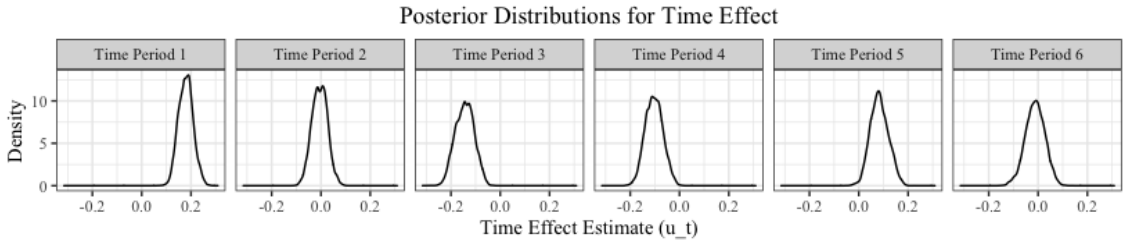

Figure 3: Posterior distributions for u_t, based on our reproduction of the code provided by the authors.

small. The estimates pulled from these plots are too close to zero when compared with the mean intercept term of -102 that we estimated. A visual inspection of the posterior distributions for the v_{tm} parameters

showed a similar situation. In short, although the spatio-temporal impacts are clearly helping the model based on its predictive power, they are broken up into too many different parameters causing the model to be unnecessarily complex.

After reproducing the work done in the paper, we turned to applying this same model to new data. Due to the prevalence of wide-spread skepticism regarding the COVID-19 vaccines, we decided to model COVID-19 vaccine efficacy in the same way using available hospitalization data. We were able to find three states in relatively close proximity with public dashboards that allowed viewing of hospitalization data by vaccination status over time. Data was gathered from Georgia [9], Virginia [10], and New York [11], with some vaccination data obtained from the USA Facts website [12]. It was compiled in the following format

region	date	vaccination_status	number_hospitalized	population	vaccination_prop
State Abbreviation	YYYY-MM-DD	Yes or No	Number per 100,000	Integer	Percentage between 0 and 1

where the particular hospitalization rate for a region is measured at the first dat of each month from May 2021 to November 2021. Patient-level data was impossible to obtain, thus our formulation ignore the covariates "sex" and "age". To replicate the model in the context of COVID-19, we utilized STAN [8] software through the rstan package in place of WinBUGS. Priors were formulated with the same distributions as described in the paper. The simulation was also run in the same way, where 10,000 iterations were computed, with the first 2,000 thrown out and only one out of every remaining 10 saved.

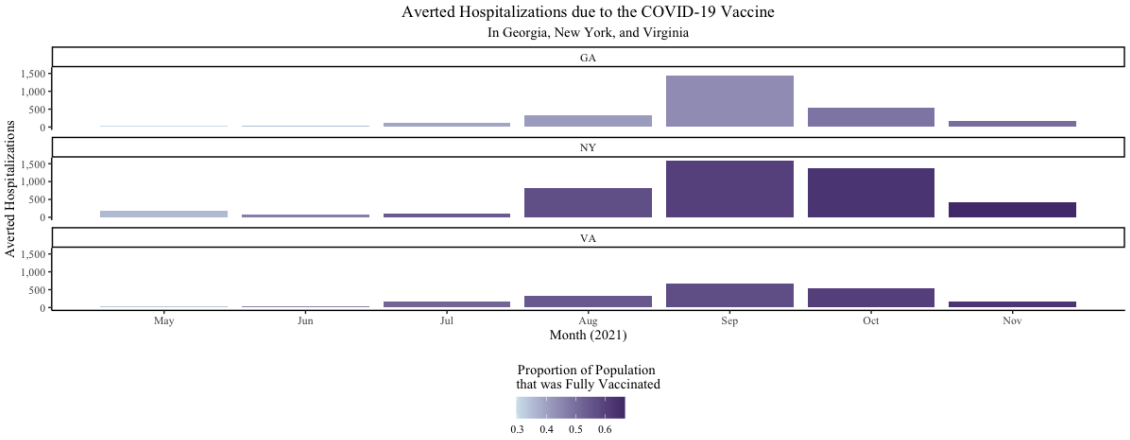

Figure 4: Hospitalizations averted due to the COVID-19 vaccine, as predicted by the model we created in the same way that the paper did.

In the end, we estimate that the COVID-19 vaccine led to 9,157 averted hospitalizations in Georgia, Virginia, and New York between May and November 2021. The distribution of these averted hospitalizations can be seen in Figure 4. We should note that we did run this process twice, and the two estimates we got for averted hospitalizations had a difference of about 2,000. Although we do not have time to run the simulation repeatedly so as to get a distribution of averted hospitalizations, we think it would be an interesting exercise.

Conclusion. Our reproduction of the paper's estimate yielded a number that was noticeably larger. After examining individual posteriors, we noticed certain distributions yielded wide variances. Intuitively, this makes sense; vaccines are not perfect and there is always a significant random effect element for certain individuals. Overall, we were pleased to obtain an estimate somewhat close to paper's results.

Our model reproduction with Covid-19 data gave interesting results. However, due to the lack of granular data, we were unable to produce a robust estimate using this model. For example, given data on which Covid-19 vaccines specific individual used, we can obtain efficacy rates for mRNA vaccines versus Johnson & Johnson shots. This is something worth exploring as new vaccines become available to the developing world.

References

[1] Mónica López-Lacort, Alejandro Orrico-Sánchez, Miguel Ángel Martínez-Beneito, Cintia Muñoz-Quiles, and Javier Díez-Domingo. Spatio-temporal impact of self-financed rotavirus vaccination on rotavirus and acute gastroenteritis hospitalisations in the Valencia region, Spain. *BMC Infectious Diseases*, 20(1):656, December 2020.

[2] Thea Kølsen Fischer, Cécile Viboud, Umesh Parashar, Mark Malek, Claudia Steiner, Roger Glass, and Lone Simonsen. Hospitalizations and Deaths from Diarrhea and Rotavirus among Children <5 Years of Age in the United States, 1993–2003. *The Journal of Infectious Diseases*, 195(8):1117–1125, April 2007.

[3] Eleanor Burnett, Christine L. Jonesteller, Jacqueline E. Tate, Catherine Yen, and Umesh D. Parashar. Global Impact of Rotavirus Vaccination on Childhood Hospitalizations and Mortality From Diarrhea. *The Journal of Infectious Diseases*, 215(11):1666–1672, June 2017.

[4] Gavin Shaddick. Implementing Models in WinBugs, October 2008.

[5] Francisca Corpas-Burgos, Gonzalo García-Donato, and Miguel A. Martinez-Beneito. Some findings on zero-inflated and hurdle poisson models for disease mapping. *Statistics in Medicine*, 37(23):3325–3337, October 2018.

[6] Leonhard Held, Niel Hens, Philip D O'Neill, and Jacco Wallinga. *Handbook of infectious disease data analysis*. CRC Press, 2019.

[7] Yingcun Xia, Ottar N Bjørnstad, and Bryan T Grenfell. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *The American Naturalist*, 164(2):267–281, 2004.

[8] Daniel Lee, Bob Carpenter, Peter Li, Mitzi Morris, Michael Betancourt, Maverickg, Marcus Brubaker, Rob Trangucci, Marco Inacio, Alp Kucukelbir, Stan Buildbot, Bgoodri, Seantalts, Jeffrey Arnold, Dustin Tran, Matt Hoffman, Charles Margossian, Martin Modrák, Avraham Adler, Krzysztof Sakrejda, Alexey Stukalov, Mike Lawrence, Rob J Goedman, Kevin S. Van Horn, Aki Vehtari, Jonah Gabry, Juan Sebastián Casallas, and Ben Bales. Stan-Dev/Stan: V2.17.1, December 2017.

[9] Georgia COVID-19 Breakthrough Report. https://breakthroughreports.s3.amazonaws.com/ Breakthrough+Report_211130.html.

[10] Virginia COVID-19 Cases by Vaccination Status. https://www.vdh.virginia.gov/coronavirus/see-the-numbers/covid-19-in-virginia/covid-19-cases-by-vaccination-status/.

[11] New York COVID-19 Breakthrough Data. https://coronavirus.health.ny.gov/covid-19-breakthrough-data.

[12] US Coronavirus vaccine tracker. https://usafacts.org/visualizations/covid-vaccine-tracker-states/, December 2021. Section: covid-19.

[1]

[1]The code used to reproduce the paper and perform our own analysis can be found in the github repository at https://github.com/srmatth/stats295-inf-diseases in the "Final Project" directory. The paper code was reproduced in the file title "STAN_paper_code.Rmd" and the original COVID-19 vaccine analysis was done in the file titled "reproduce_w_covid_data.Rmd". The actual model reproduction result (.Rdata) object is not in the repository because it exceeds the Github size limits.