# Estimating COVID-19 Incidence During the Pre-Vaccination Era in North Carolina Counties by Leveraging Spatial Characteristics

Cindy Pang
Faculty Advisor: Paul Delamater (UNC Department of Geography)

## Background

- Testing availability during the COVID-19 pandemic has made it difficult to accurately estimate the true number of infections, leading to an underestimate of disease prevalence and basic reproductive number
- Testing limitations in the early stages of the pandemic resulted in capturing only a pool of mainly symptomatic persons, whereas many infected individuals are asymptomatic
- Age plays a significant role in asymptomatic infections, with the highest percentage of asymptomatic infections found in persons 20 years old or younger.
- Geographical disparities and lack of access to testing can further complicate estimating COVID-19 infections.

## Research Objectives

- Develop a model that estimates COVID-19 infection incidence for North Carolina Counties from March 2020 to January 1, 2021
- Address the challenge of limited data sources at the county-level (only reported deaths available)
- Utilize Age-Specific Infection Fatality Ratios (IFR) for each county to reconstruct the infection curve

## Methods

COVID-19 incidence will be estimated using the back-casting method outlined in Phipps, Grafton, and Kompas. Assume that the time from infection to death ($t_d$) follows a Gamma distribution with parameters $\alpha = \left(\frac{\mu}{s}\right)^2$ and $\beta = \frac{s^2}{\mu}$ where $\mu$ is the time from infection to death and $s$ is the standard deviation. The number of new infections ($n_i(t', t)$) estimated to occur on day $t'$ resulting in fatalities on day $t$ is shown in equation 1:

*(Eqn. 1)*

$$n_i(t', t) = \frac{N_f(t) * f(t - t'; \alpha, \beta)}{IFR}$$

In equation 1, $N_f(t)$ is the number of new fatalities that occur on day $t$ and $f(t - t'; \alpha, \beta)$ is the probability density function (pdf) of the Gamma distribution. *IFR* is the infection-fatality ratio.

In our analysis, we will split the IFR by age-group for the hierarchical model and aggregate the total number of COVID-19 incidence by county.

Finally, the estimated total number of new infections ($N_i(t)$) on day $t'$ is shown in equation 6 where $F(t_0 - t'; \alpha, \beta)$ is the cumulative density function of the Gamma distribution and $t_0$ is the most recent day for which fatality statistics are available.

*(Eqn. 2)*

$$N_i(t') = \frac{1}{F(t_0 - t'; \alpha, \beta)} \sum_{t=t'+1}^{t_0} n_i(t', t)$$

Error bounds were obtained via Monte-Carlo methods which were used to sample parameter uncertainty. 10,000 MCMC simulations were generated with random draws from within the uncertainty range for each parameter using a Gaussian probability distribution.

## Results

Two ensemble models will be used as comparisons:

**(1) Case Multiplier Model:** From the NCDHHS Dataset, the number of reported positive cases were multiplied by a factor of four[5], which represents the ratio between infections and positive cases in the United States from February 2020 and September 2021.

**(2) Death to Infection Estimate:** The number of reported deaths was divided by the infection fatality ratio (0.0065) found by Reese et al. [5] and lagged the infection by the median number of days from death to reporting (7 days for ages 18-64 and 6 days for ages 65 and above).

Other comparisons models will include **(3) Reported Weekly Cases from NCDHHS and (4) Ensemble Average, which is the average of (1), (2), and (3).**

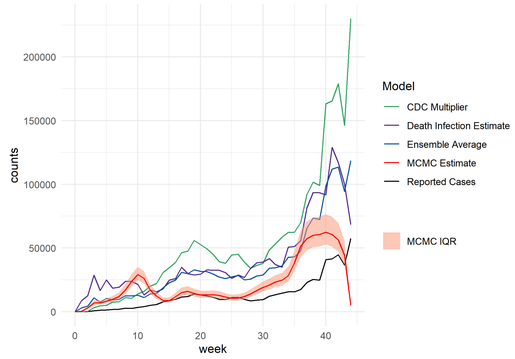


Figure 1. *Time Series of Daily Infections in North Carolina*

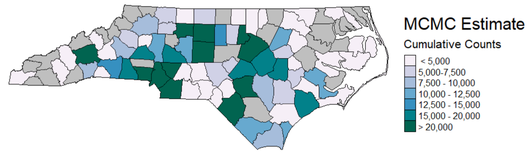


Figure 2. *Distribution of Cumulative Infections under MCMC Model across the State of NC*

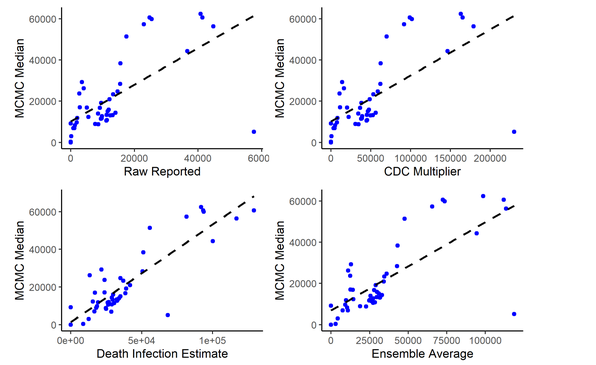


Figure 3. *Scatterplot Comparison of Temporal Similarity between MCMC Estimate and Ensemble Models*

| Model | Correlation Coefficient (ρ) | p-value |
|---|---|---|
| Reported cases | 0.669 | < 0.001 |
| CDC Multiplier | 0.669 | < 0.001 |
| Death Infection Estimate | 0.692 | < 0.001 |
| Ensemble Average | 0.689 | < 0.001 |

Table 1. *Spearman's Rank Correlation Test for Temporal Similarity*

In the statewide model, the Root Mean Square Error was the lowest between the MCMC Estimates and Reported Cases (RMSE = 16,251.23) and was the highest for the CDC Multiplier (RMSE = 51,588.05). The RMSE values for the Death Infection Estimate and Ensemble Average are 24,332.68 and 24,761.28, respectively. Despite the RMSE being the lowest between the MCMC Estimates and Reported Cases, the magnitude of error is still large ($10^4$), indicating that the Reported Cases did not capture many of the true infections.

## Discussion

1. MCMC model shows mixed performance across temporal similarity and magnitude of error tests.
2. MCMC model exhibits strong temporal similarity with the Death Infection Estimates model.
3. MCMC model is not robust when accounting for the magnitude of error, with high RMSE values across all ensemble models.
4. Some North Carolina counties had more robust models than others, with the MCMC model bounded by the Reported Cases and the Death Infection Estimate Model.
5. MCMC model is limited by missing death data and the time between when a person died and when the death was reported, which may affect temporal aspects of the epidemic curve.

References:

- Irons NJ, Raftery AE. Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proc Natl Acad Sci USA*. 2021;118(31):e2103272118. doi:10.1073/pnas.2103272118
- National Academies of Sciences, Engineering, and Medicine. *Evaluating Data Types: A Guide for Decision Makers Using Data to Understand the Extent and Spread of COVID-19*. National Academies Press; 2020:25826. doi:10.17226/25826
- Phipps SJ, Grafton RQ, Kompas T. Robust estimates of the true (population) infection rate for COVID-19: a backcasting approach. *R Soc open sci*. 2020;7(11):200909. doi:10.1098/rsos.200909
- Variation in the COVID-19 infection–fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. *The Lancet*. 2022;399(10334):1469-1488. doi:10.1016/S0140-6736(21)02867-1
- Reese H, Iuliano AD, Patel NN, et al. Estimated Incidence of Coronavirus Disease 2019 (COVID-19) Illness and Hospitalization—United States, February–September 2020. *Clinical Infectious Diseases*. 2021;72(12):e1010-e1017. doi:10.1093/cid/ciaa1780
- Estimated COVID-19 Burden | CDC. Accessed March 22, 2023. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html
- How to Use our Data. Johns Hopkins Coronavirus Resource Center. Accessed December 4, 2022. https://coronavirus.jhu.edu/about/how-to-use-our-data
- NCDHHS_COVID-19_DataDownload_STORY | Tableau Public. Accessed May 31, 2022. https://public.tableau.com/app/profile/ncdhhs.covid19data/viz/NCDHHS_COVID-19_DataDownload_Story_16220681778000/DataBehindtheDashboards