Bayesian Inference with PyBioNetFit of State-Level R_0 values for COVID-19 across the US

Abhishek Mallela[1], Jacob Neumann[2], Ely F. Miller[3], Ye Chen[4], Richard G. Posner[3], Yen Ting Lin[5], and William S. Hlavacek[1]

***Short Abstract* — Disease transmission dynamics in a given population are often characterized by calculating or estimating the population-specific basic reproduction number R_0. Using Bayesian inference and the PyBioNetFit software package, we estimated R_0 for COVID-19 in each of the 50 states using 2020 surveillance data (i.e., daily number of cases detected). Our estimates indicate that disease transmission varied considerably across the states and identify regions at risk of relatively high rates of disease transmission if a new disease similar to COVID-19 emerges. Our analysis demonstrates that PyBioNetFit is a practical tool for solving Bayesian inference problems.**

***Keywords* — mathematical model; coronavirus disease 2019 (COVID-19); basic reproduction number; Bayesian inference; PyBioNetFit**

I. INTRODUCTION

It is well-known that the rate of COVID-19 disease transmission depends not only on virological features, but also on those of the exposed population, including biological, sociobehavioral, and environmental factors (e.g., age structure, commuting patterns, and architecture) [1]. COVID-19 surveillance data from 2020 offers an unprecedented opportunity to identify regional populations within the US that are susceptible to rapid spread of an aerosol-transmitted virus.

With this information, population features that correlate with susceptibility to relatively fast disease transmission can also be identified. Using PyBioNetFit [2], we focused in this work on addressing questions about why COVID-19 and similar diseases spread more rapidly in some populations/regions than in others.

II. RESULTS

In the US, disease surveillance data are collected by county health departments. During the COVID-19 pandemic, analysts most commonly aggregated county-level data to characterize disease transmission within states, presumably because state governors drove COVID-19 responses [3].

How did early rates of COVID-19 transmission vary across the 50 states of the US? In our work reported in *Viruses* in 2022 [4], we addressed this question by parameterizing a compartmental model for COVID-19 transmission dynamics for each of the 50 states for consistency with 2020 surveillance data. This model was similar to the model that we developed in 2020 [5]. We used a model parameterization approach based on Bayesian inference and a practical Markov chain Monte Carlo sampling algorithm in PyBioNetFit [6].

From parameterized models, we obtained estimates of the basic reproduction number R_0, the expected number of secondary cases arising from an index case, using the next-generation matrix approach [7]. In other words, we estimated R_0 for each state. Maximum a posteriori (MAP) estimates for R_0 range from 7.1 for New Jersey to 2.3 for Wyoming, indicating that disease transmission varies considerably across states.

III. CONCLUSION

Our findings imply that the estimation of R_0 for US subpopulations will help to identify populations that are susceptible to rapid transmission of respiratory diseases similar to COVID-19. In addition, we demonstrated that PyBioNetFit can be used to successfully solve real-world Bayesian inference problems. These results should help the nation prepare for future pandemics.

REFERENCES

[1] Delamater PL, et al. (2019) Complexity of the Basic Reproduction Number (R_0). *Emerg Infect Dis* **25**, 1-4.

[2] Mitra ED, et al. (2019) PyBioNetFit and the Biological Property Specification Language. *iScience* **19**, 1012-1036.

[3] Weissert CS, et al. (2021) Governors in control: Executive orders, state-local preemption, and the COVID-19 pandemic. *Publius* **51**, 396-428.

[4] Mallela A, et al. (2022) Bayesian Inference of State-Level COVID-19 Basic Reproduction Numbers across the United States. *Viruses* **14**, 157.

[5] Lin YT, et al. (2021) Daily Forecasting of Regional Epidemics of Coronavirus Disease with Bayesian Uncertainty Quantification. *Emerg Infect Dis* **27**, 767-778.

[6] Neumann J, et al. (2022) Implementation of a practical Markov chain Monte Carlo sampling algorithm in PyBioNetFit. *Bioinformatics* **38**, 1770-1772.

[7] Diekmann A, Heesterbeek JA, Roberts MG (2010) The construction of next-generation matrices for compartmental epidemic models. *J R Soc Interface*. **7**, 873–875

[1]Los Alamos National Laboratory, Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos, NM 87545, USA.

[2]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA.

[3]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA.

[4]Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ 86011, USA.

[5]Los Alamos National Laboratory, Information Sciences Group, Computer, Computational and Statistical Sciences Division, Los Alamos, NM 87545, USA.