Hospitals as Reservoirs for Clostridioides difficile in the Community

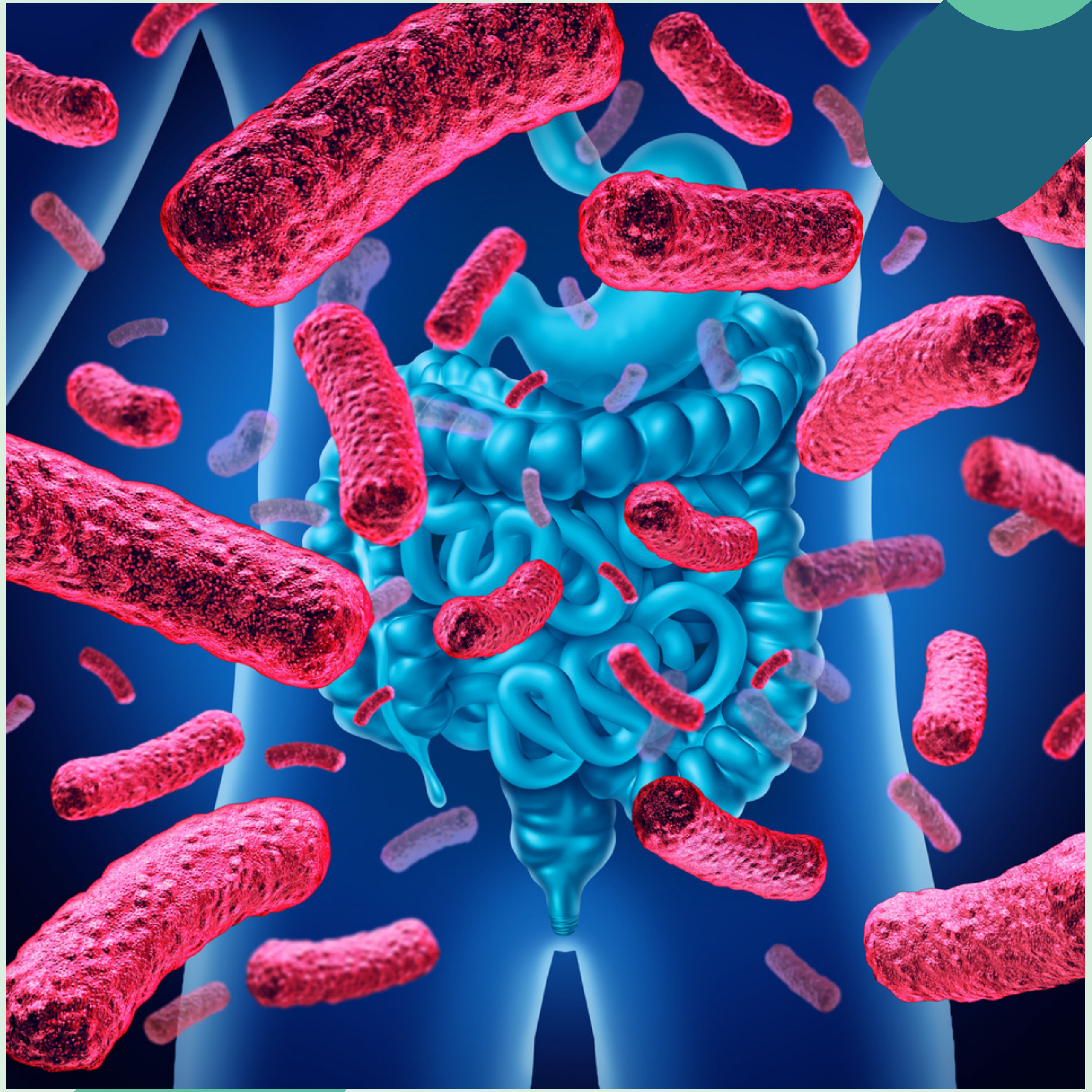
Setara Nusratty, Sofía Rodríguez-Carrió, Quinn Stoddard-O'Neill

Daniel K. Sewell, Ph.D.

What is Clostridioides difficile (C. difficile)?

What is C. difficile?

- A bacterium that causes infection in the large intestine.
- One of the most common healthcare facility-acquired infections.
- Symptoms: diarrhea to life-threatening colitis.
- Transmission occurs by the fecal-oral route.
- Spores could persist in the environment for a long time.
- Asymptomatic patients could still spread the infection.



https://www.health.harvard.edu/blog/stool-transplants-are-now-standard-of-care-for-recurrent-c-difficile-infections-2019050916576

What is C. difficile?

- Risk factors:
  - Presence of other CDI patients in the hospital
  - Use of antibiotics
  - Old age
  - Longer hospital stays
  - Comorbidities
  - Exposure to agents that reduce levels of gastric acid

https://therivernewsroom.com/coronavirus-roundup-state-shifts-data-focus-hospitalizations-4000/

OBJECTIVE

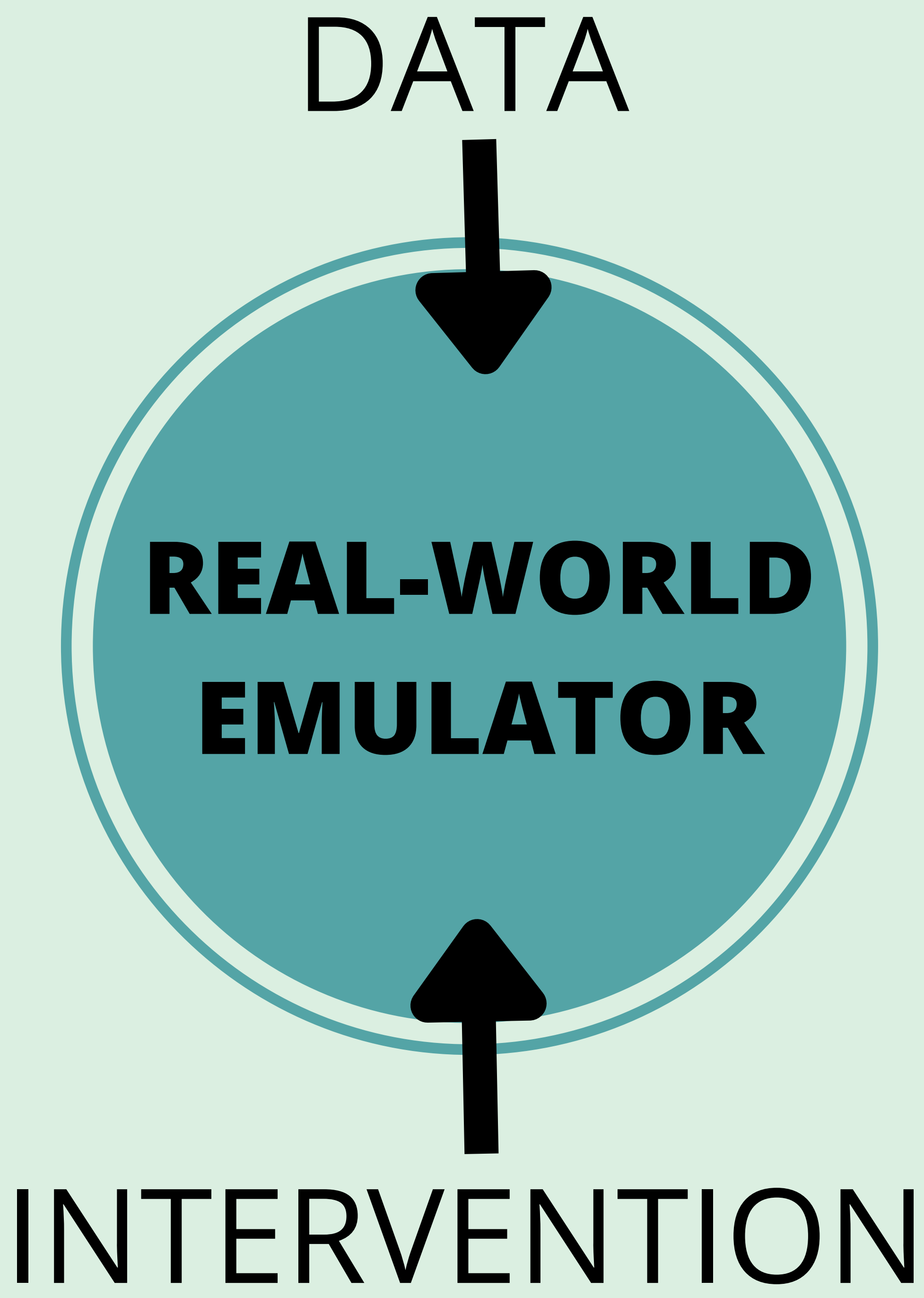
- To determine if individuals that experience healthcare facility-onset C. difficile infections (HCFO-CDIs) are contributing to the spread of community-onset C. difficile infections (CO-CDIs).

STUDY DESIGN

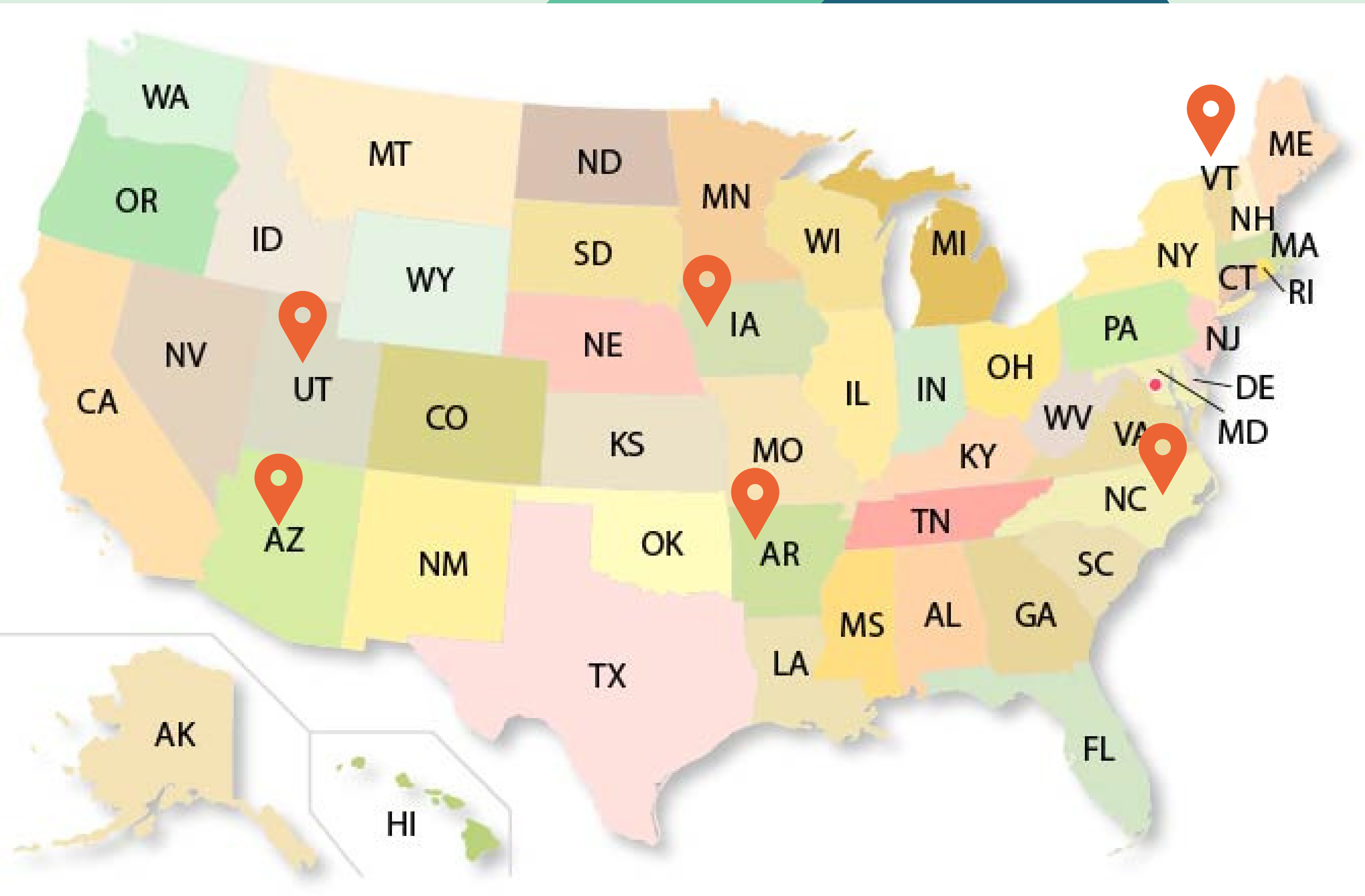
- Data cleanup
- Create a real-world emulator to predict CO-CDI cases
- Create a simulation that predicts state counties and their neighboring counties' CO-CDI cases based on previous months

INTERVENTION

- Made a counterfactual scenario of negating the effect of HCFO-CDIs on community spread.
- Manage to predict the CO-CDI cases if the detrimental effect of HCFO-CDIs was eliminated.

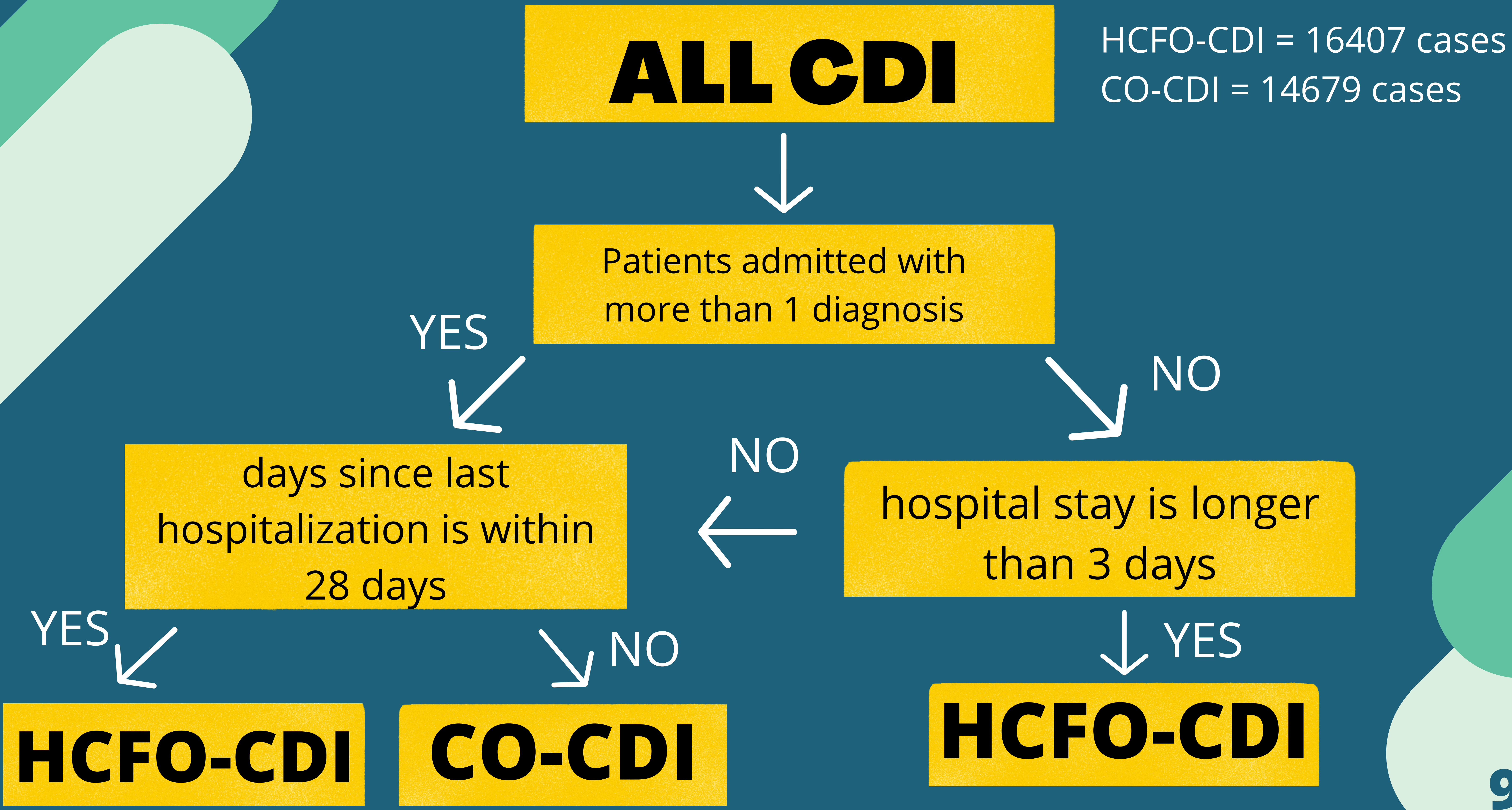


DATA



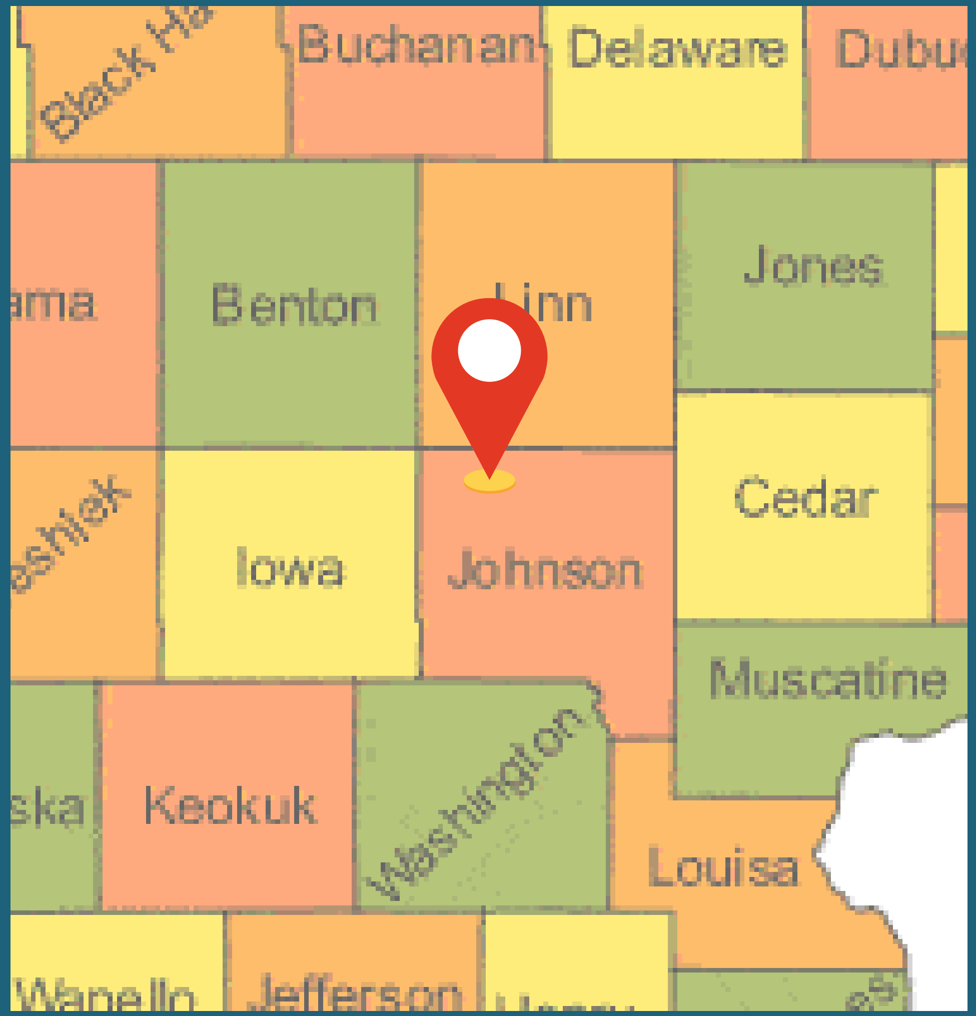
https://www.nationsonline.org/oneworld/us_states_maps.htm

- State Inpatient Database from the Healthcare Cost and Utilization Project (HCUP SID)
- Over 20 million hospital inpatient visits from six U.S. states (VT, NC, AZ, AR, UT, IA) from 2003 to 2015
- Census data
- County adjacency spatial data to predict local spread


ALL CDI
HCFO-CDI = 16407 cases
CO-CDI = 14679 cases
Patients admitted with more than 1 diagnosis
YES
NO
days since last hospitalization is within 28 days
NO
hospital stay is longer than 3 days
YES
NO
YES
HCFO-CDI
CO-CDI
HCFO-CDI

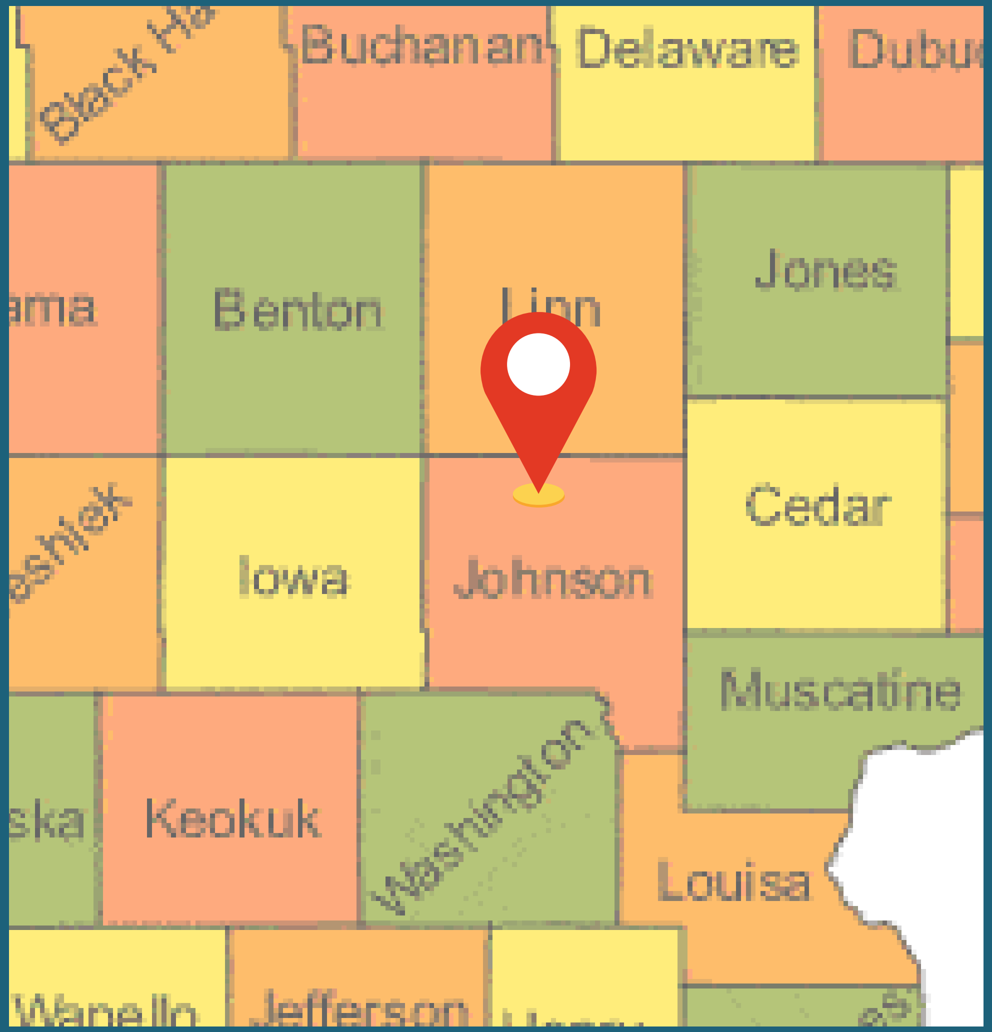
Temporal

PAST



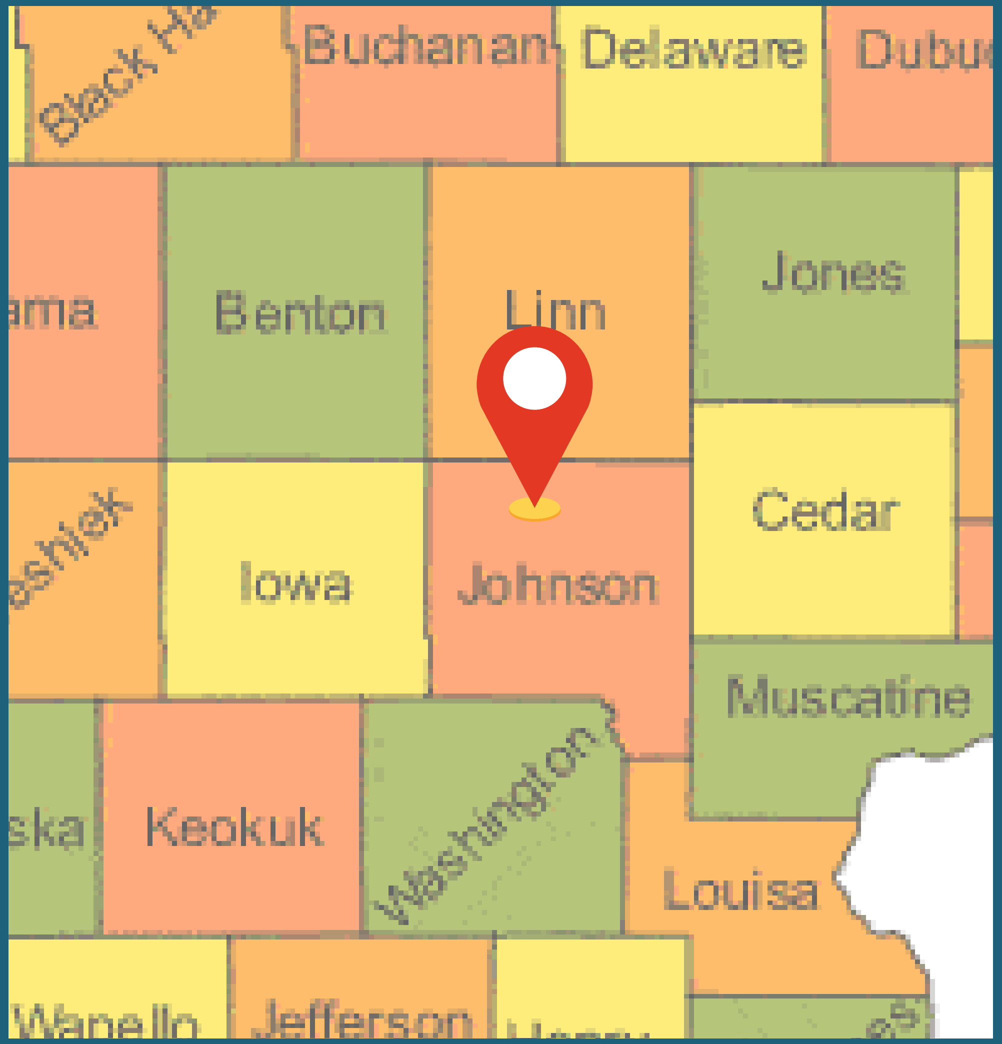
JUNE

PRESENT



JULY

FUTURE

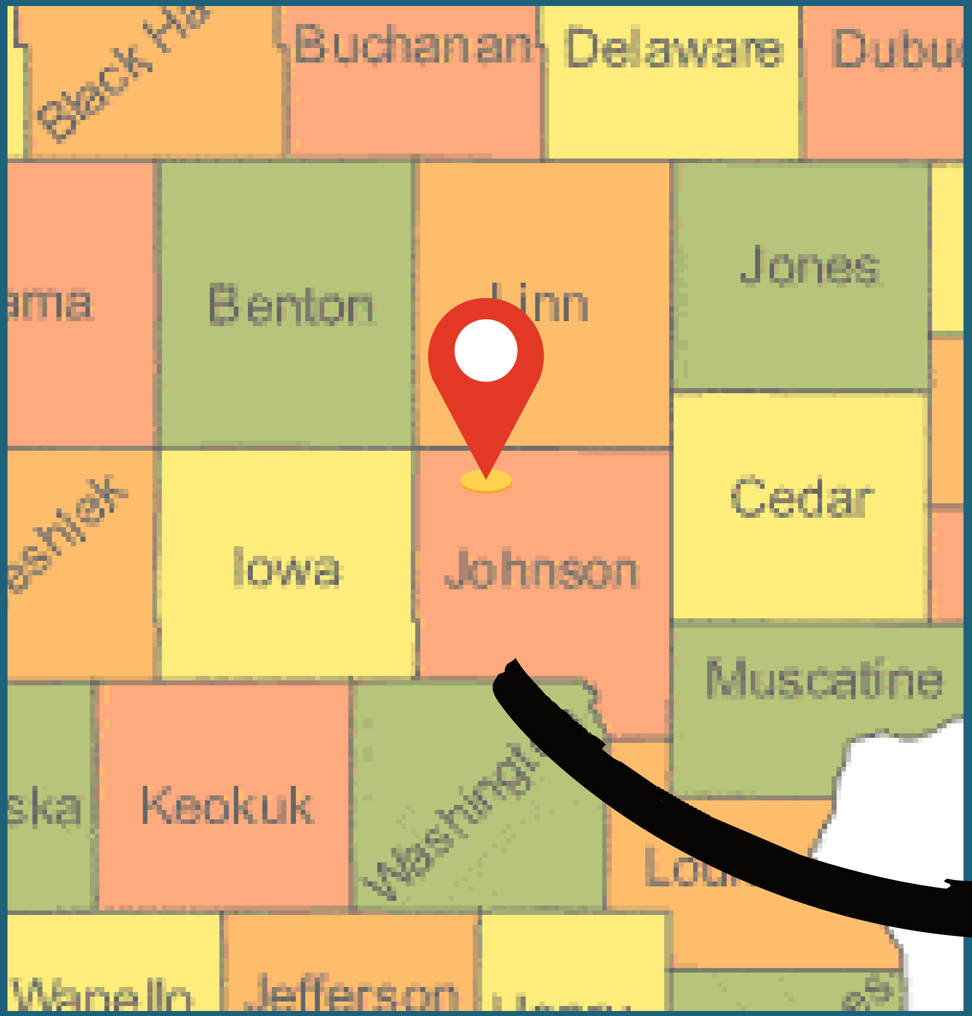


AUGUST

https://geology.com/state-map/iowa.shtml

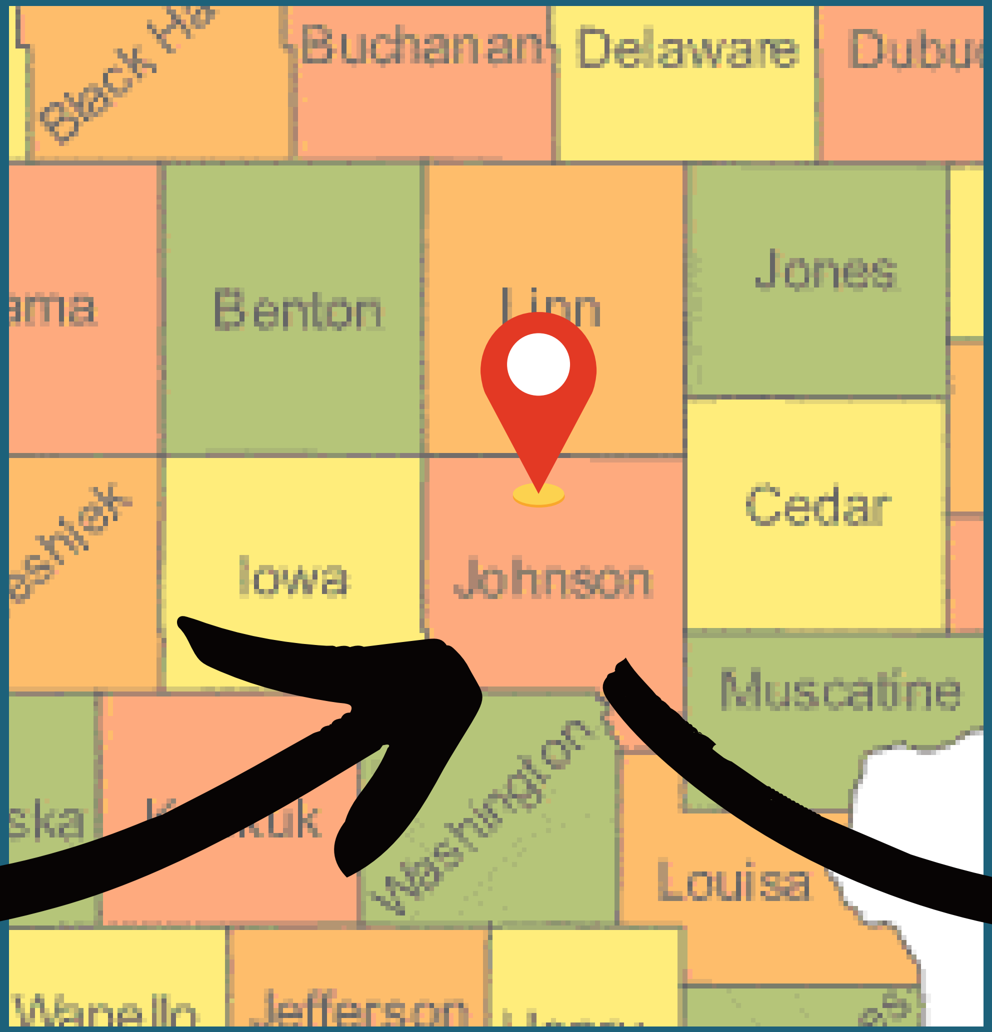
Temporal

PAST



JUNE

PRESENT



JULY

FUTURE

AUGUST

https://geology.com/state-map/iowa.shtml

Spatial

PAST

JUNE

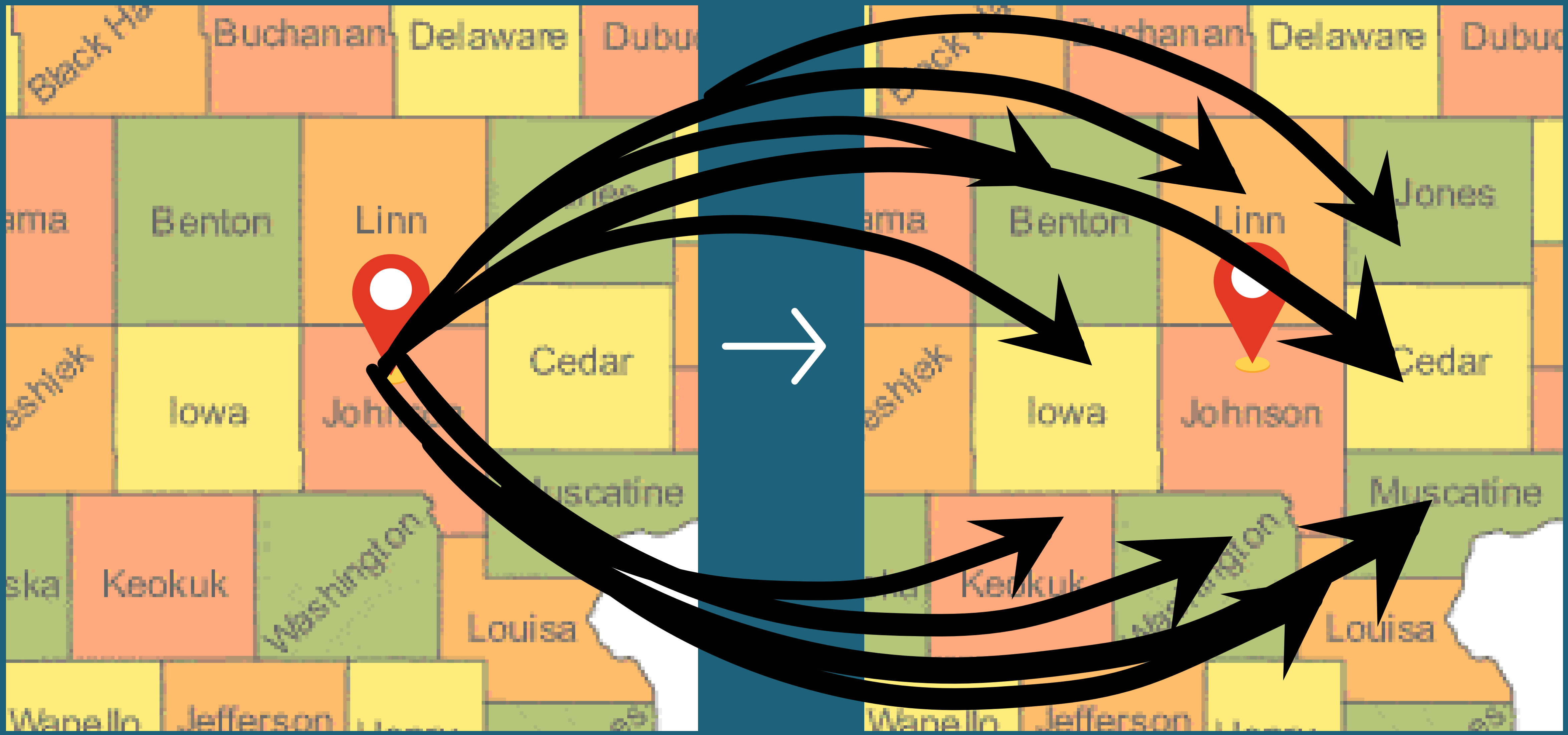
PRESENT

JULY

https://geology.com/state-map/iowa.shtml

Spatial

PAST **PRESENT**

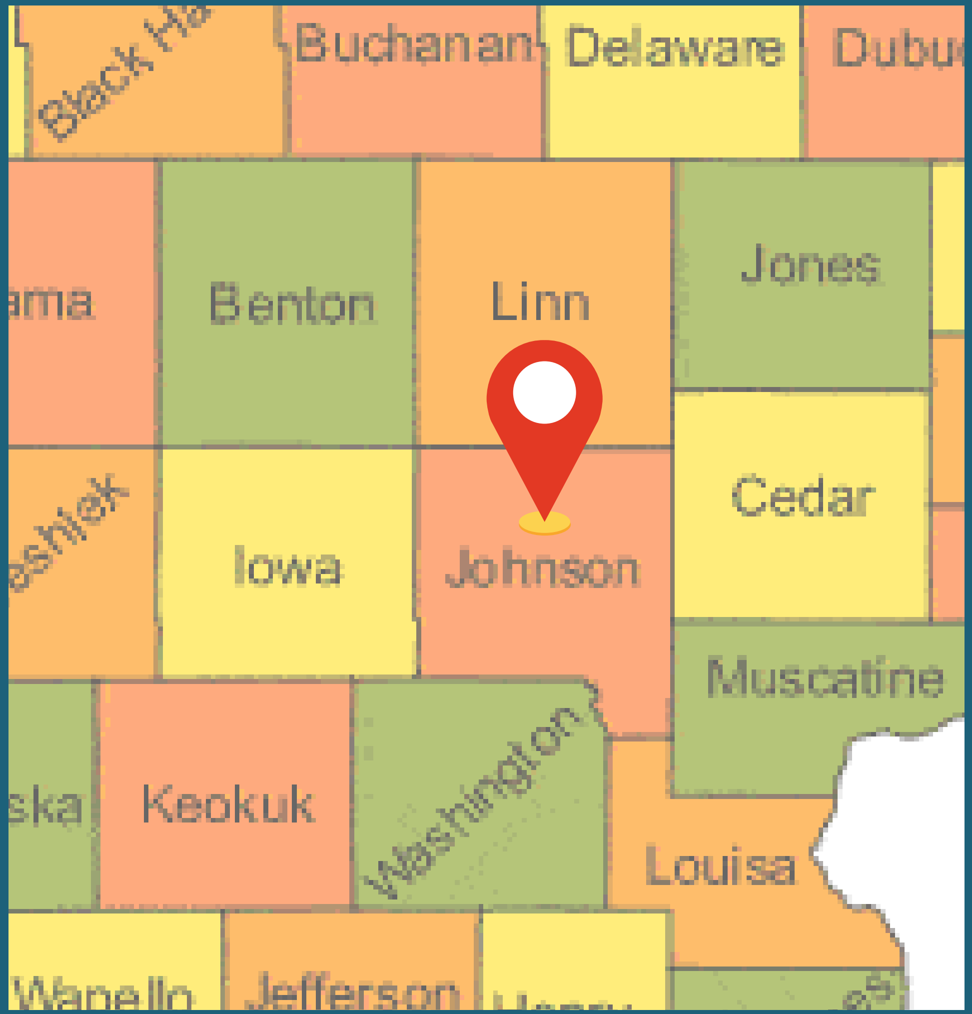


JUNE JULY

https://geology.com/state-map/iowa.shtml

SIMULATION VISION

PRESENT



JULY

- NO HCFO-CDIs

FUTURE

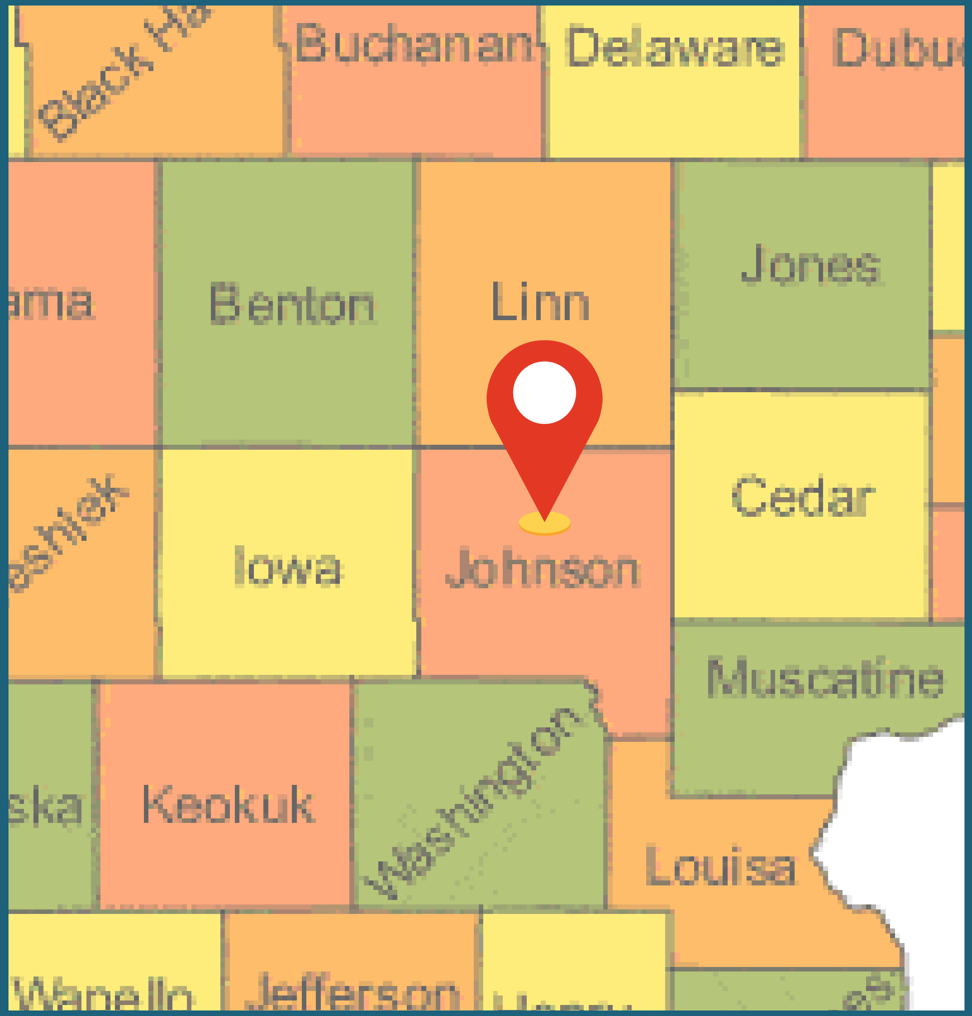
AUGUST

- NO HCFO-CDIS
- CO-CDIs start decreasing

https://geology.com/state-map/iowa.shtml

SIMULATION VISION

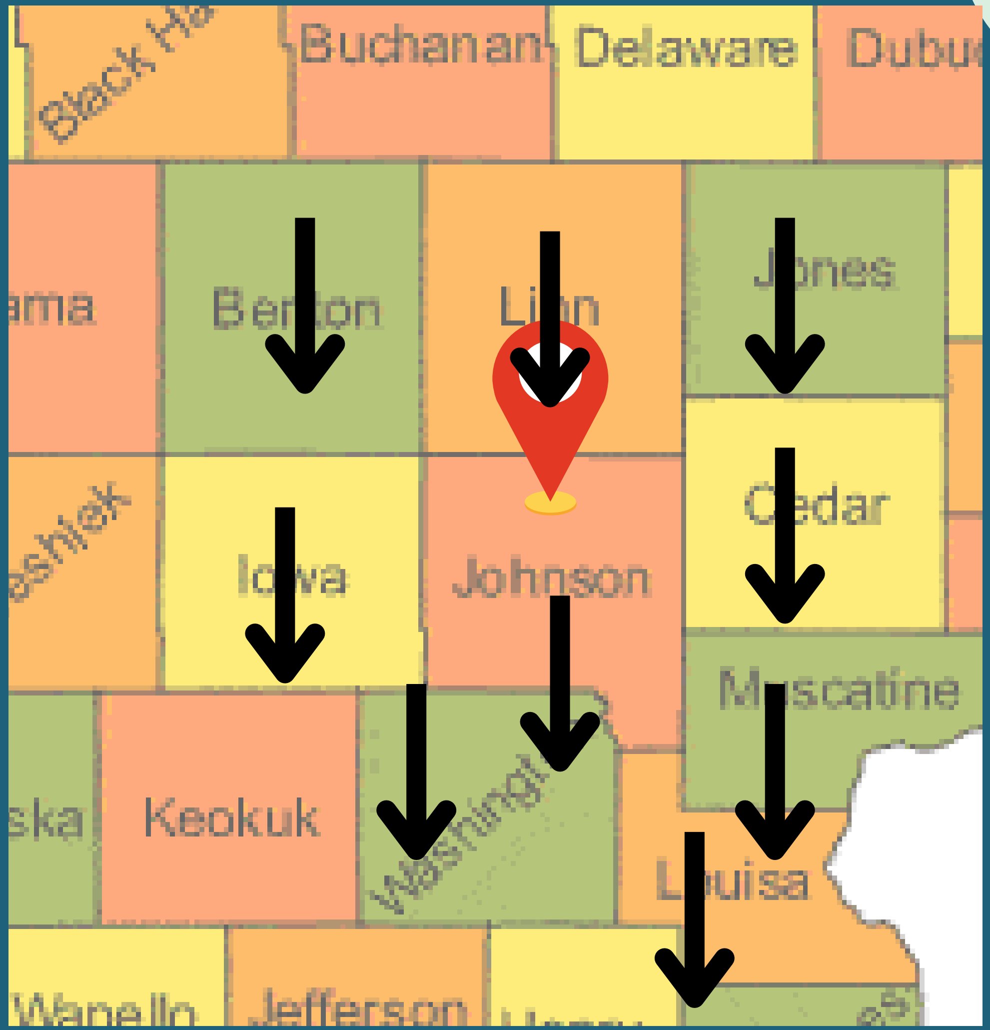
PRESENT



JULY

- NO HCFO-CDIs

FUTURE



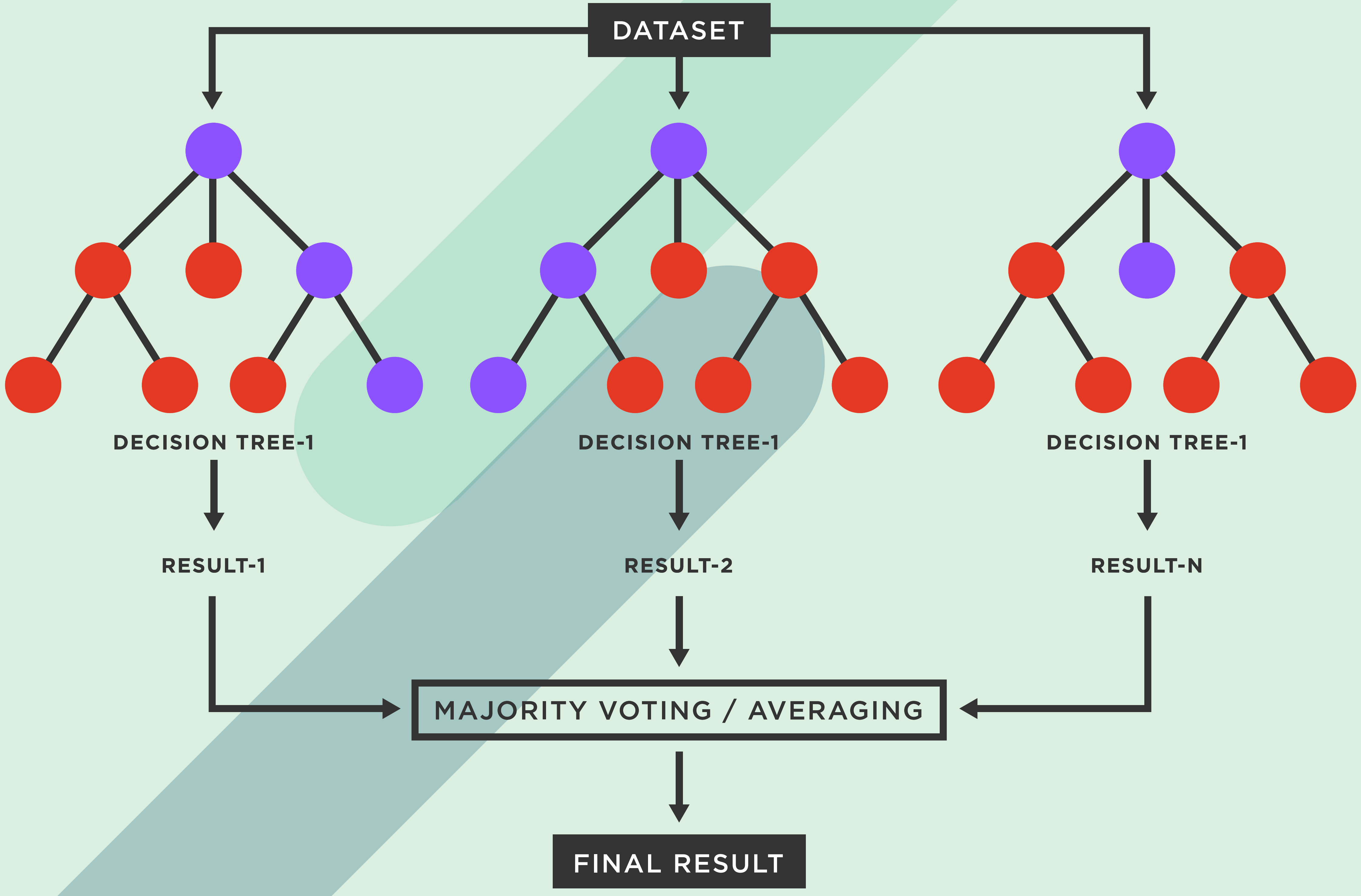
AUGUST

- NO HCFO-CDIS
- CO-CDIs start decreasing

https://geology.com/state-map/iowa.shtml

METHOD: RANDOM FOREST

- Modification of bagging
- Builds numerous de-correlated regression trees to predict an outcome
- Split data set into "train" (80%) and "test" (20%)
- Use the training data to build the random forest and predict the test data



https://www.tibco.com/reference-center/what-is-a-random-forest

RANDOM FOREST

- Predictor variables:
  - Year
  - State
  - Seasonality
  - Population 65 and older
  - Population under 65
  - HCFO-CDIs
  - HCFO-CDIs from the month prior
  - HCFO-CDIs in neighboring counties
  - CO-CDIs from the month prior
  - CO-CDIs in neighboring counties
- Response variable:
  - CO-CDIs

RANDOM FOREST

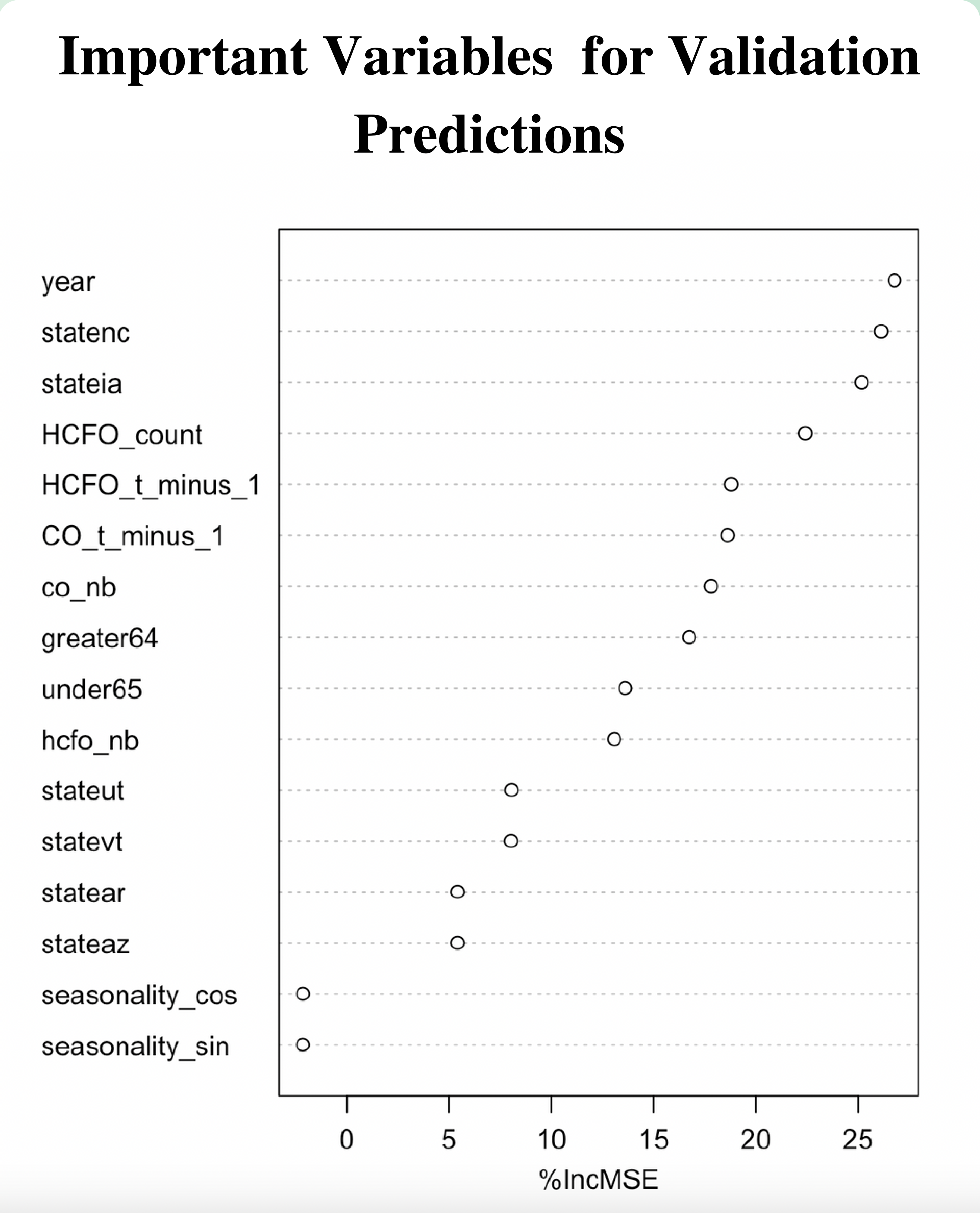
- Tried several combinations of random forest tuning parameters
- Calculated the lower mean squared error (MSE) value.
- Use result for final fit:
  - Mean of squared residuals: 2.214183
  - %Var explained: 92.88

Random Forest MSE Values

	Nodesize		
Mtry	5	10	15
2	2.364351	2.412077	2.424765
3	2.136656	2.137867	2.200545
4	2.113935	2.097265	2.129924
5	2.117038	2.104251	2.084148

- Implementing Random Forest without HCFO variables:
  - Mean of squared residuals: 2.310533
  - % Var explained: 92.57

RESULTS: ASSESSING CO-CDI PREDICTORS

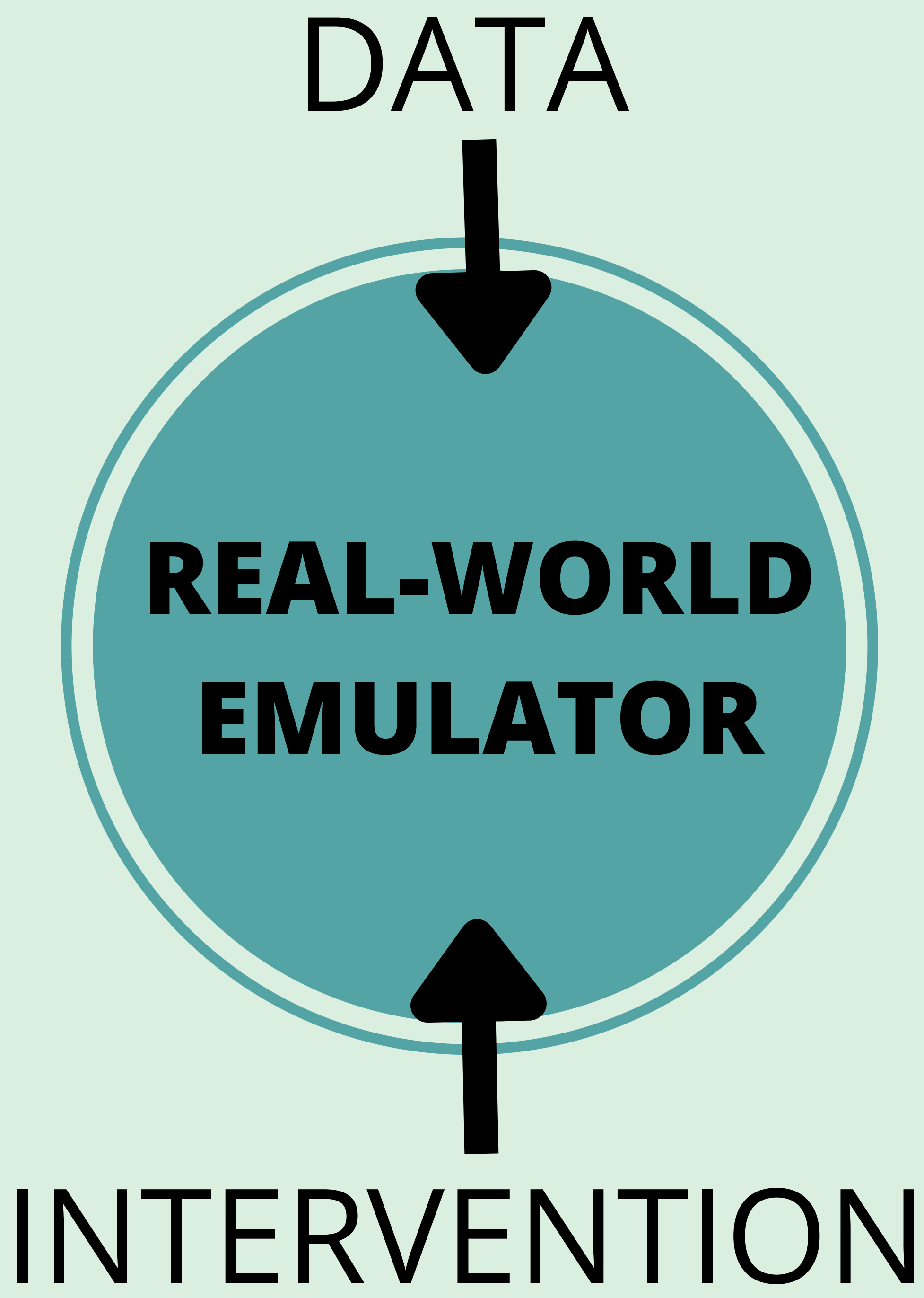


SIMULATIONS

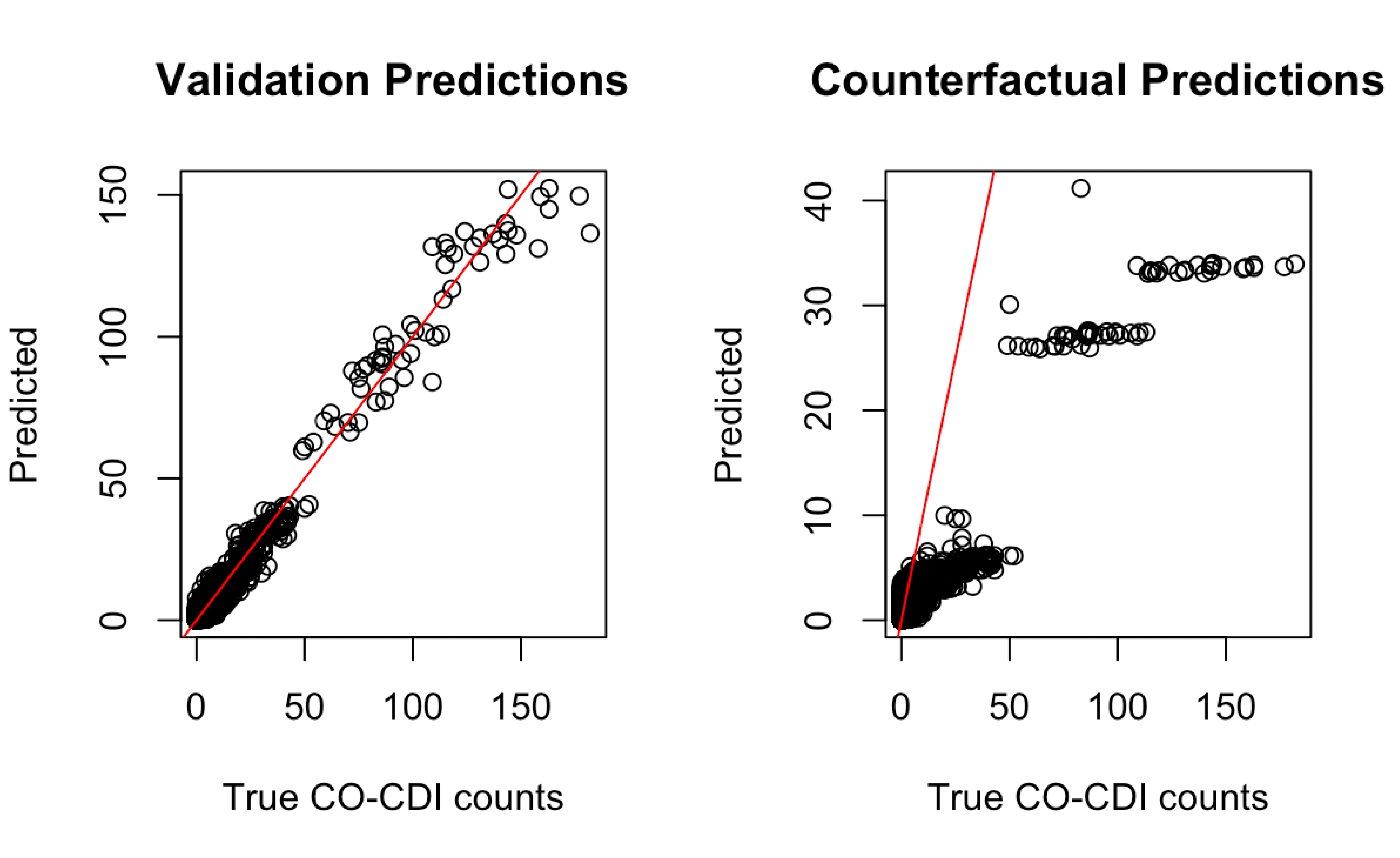
- Using the final fit from the random forest, we predict the CO-CDI cases in future months.
- We took into consideration the fips region, county adjacency, and the present and future month and year variables.

INTERVENTION

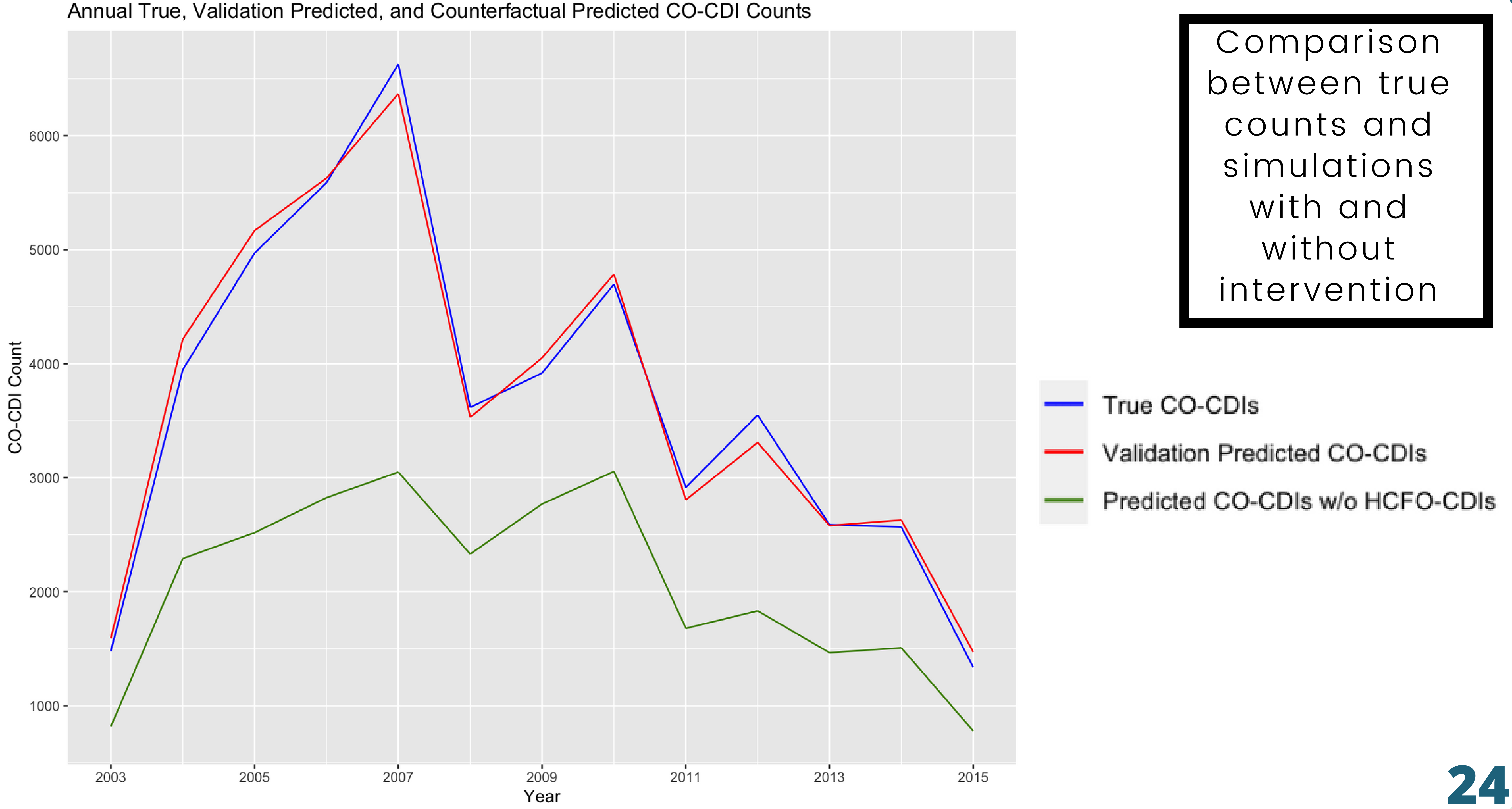
- Made a counterfactual scenario of negating the effect of HCFO-CDIs on community spread.
- Manage to predict the CO-CDI cases if the detrimental effect of HCFO-CDIs was eliminated.



RESULTS: ASSESSING THE PREDICTION OF CO-CDIS

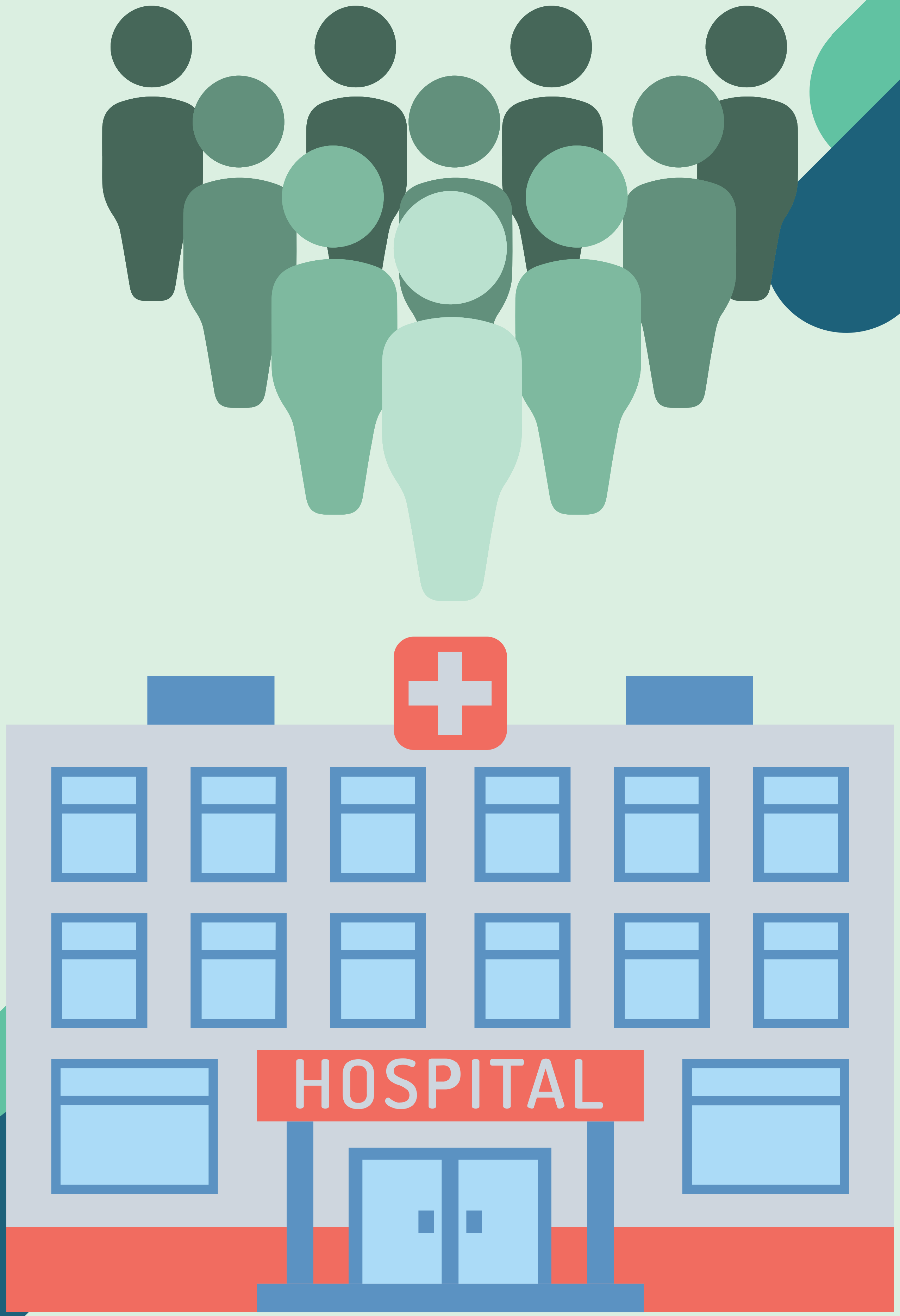


GRAPH: TOTAL ANNUAL PREDICTIONS



Comparison between true counts and simulations with and without intervention

CONCLUSION

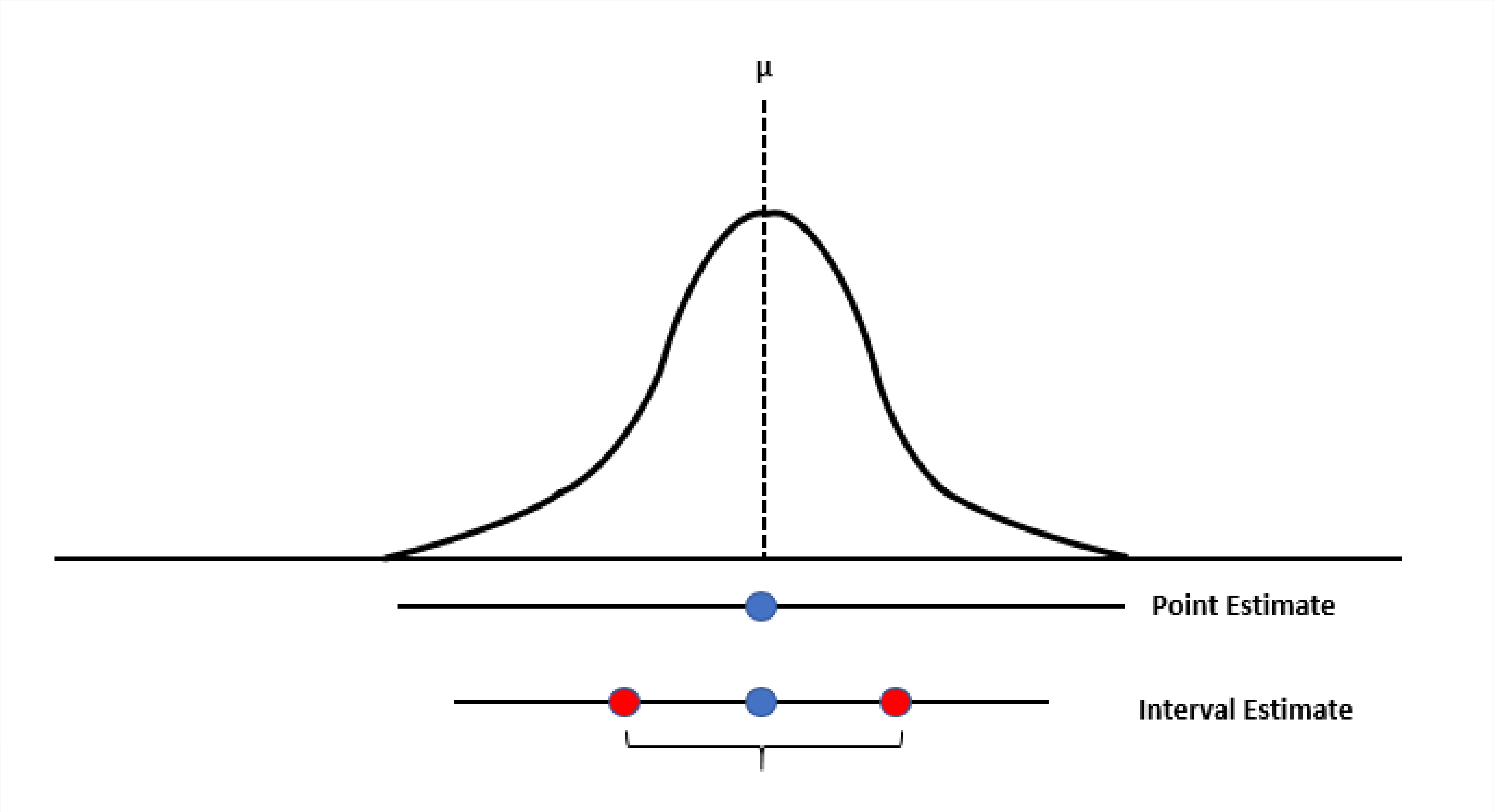


- More error in predictions omitting the HCFO variables.
- The year, number of CO-CDIs in surrounding counties, number of CDIs from the previous month, patient's age, and HCFO count are the most significant predictors of the number of CO-CDIs
- Without HCFO-CDIs, there would be a decrease in the CO-CDI cases

LIMITATIONS

- The database was composed of only 6 states.
- CDI cases that don't lead to hospitalization are not captured in the data.
- With claims data, we do not know the diagnostic errors involved in diagnosing CDI.

LIMITATIONS



- Could only compute the number of positive CDI tests.
- Obtained a point, rather than an interval, estimate.
- Took into consideration patients who were admitted to hospitals in the state in which they reside.

What's next?

- Get interval estimates from our point estimate predictions
- Implement the use of a *Poisson* distribution
- Determine another way to analyze the data
- Expand the investigation to other states and years
- Investigate CO-CDIs and socioeconomic status based on FIPS region

ACKNOWLEDGEMENTS

- Dr. Dan Sewell, Ph.D. (Mentor)
- Dr. Gideon Zamba, Ph.D. (ISIB 2022 Director)
- Eldon Sorensen (ISIB 2022 TA)
- Terry Kirk (Graduate Program Administrator)
- NHLBI Grant # HL-147231 (ISIB sponsor)

NIH National Heart, Lung, and Blood Institute

https://www.nhlbi.nih.gov/

REFERENCES

- Czepiel, J., Dróżdż, M., Pituch, H. *et al. Clostridium difficile* infection: review. *Eur J Clin Microbiol Infect Dis* **38,** 1211–1221 (2019). https://doi.org/10.1007/s10096-019-03539-6
- Mayo Foundation for Medical Education and Research. (2021, August 27). *C. difficile infection*. Mayo Clinic. Retrieved July 19, 2022, from https://www.mayoclinic.org/diseases-conditions/c-difficile/symptoms-causes/syc-20351691
- Miller, A. C., Arakkal, A. T., Sewell, D. K., Segre, A. M., Pemmaraju, S. V., & Polgreen, P. M (2022). Risk for Asymptomatic Household Transmission of Clostridioides difficile Infection Associated with Recently Hospitalized Family Members. Emerging Infectious Diseases, 28(5), 932–939. https://doi.org/10.3201/eid2805.212023
- Miller, A. C., Segre, A. M., Pemmeraju, S. V., Sewell, D. K., & Polgreen, P. M. (2020). Association of Household Exposure to Primary Clostridioides difficile Infection with Secondary Infection in Family Members. JAMA Health Forum, 3(6), e208925–e208925. https://doi.org/10.1001/jamanetworkopen.2020.8925
- *Random forests.* Random Forests · UC Business Analytics R Programming Guide. (n.d.). Retrieved July 19, 2022, from https://uc-r.github.io/random_forests
- *Regression trees.* Regression Trees · UC Business Analytics R Programming Guide. (n.d.). Retrieved July 19, 2022, from https://uc-r.github.io/regression_trees
- Sewell, D. K., Simmering, J. E., Justice, S., Pemmaraju, S. V., Segre, A. M., & Polgreen, P. M (2019). Estimating the Attributable Disease Burden and Effects of Interhospital Patient Sharing on Clostridium difficile Infections. Infection Control and Hospital Epidemiology, 40(6), 656–661. https://doi.org/10.1017/ice.2019.73

QUESTIONS?