Modeling Tick-Borne Disease in Iowa

Logan Harris, Paul Nguyen, Micaela Richter

Grant Brown, PhD

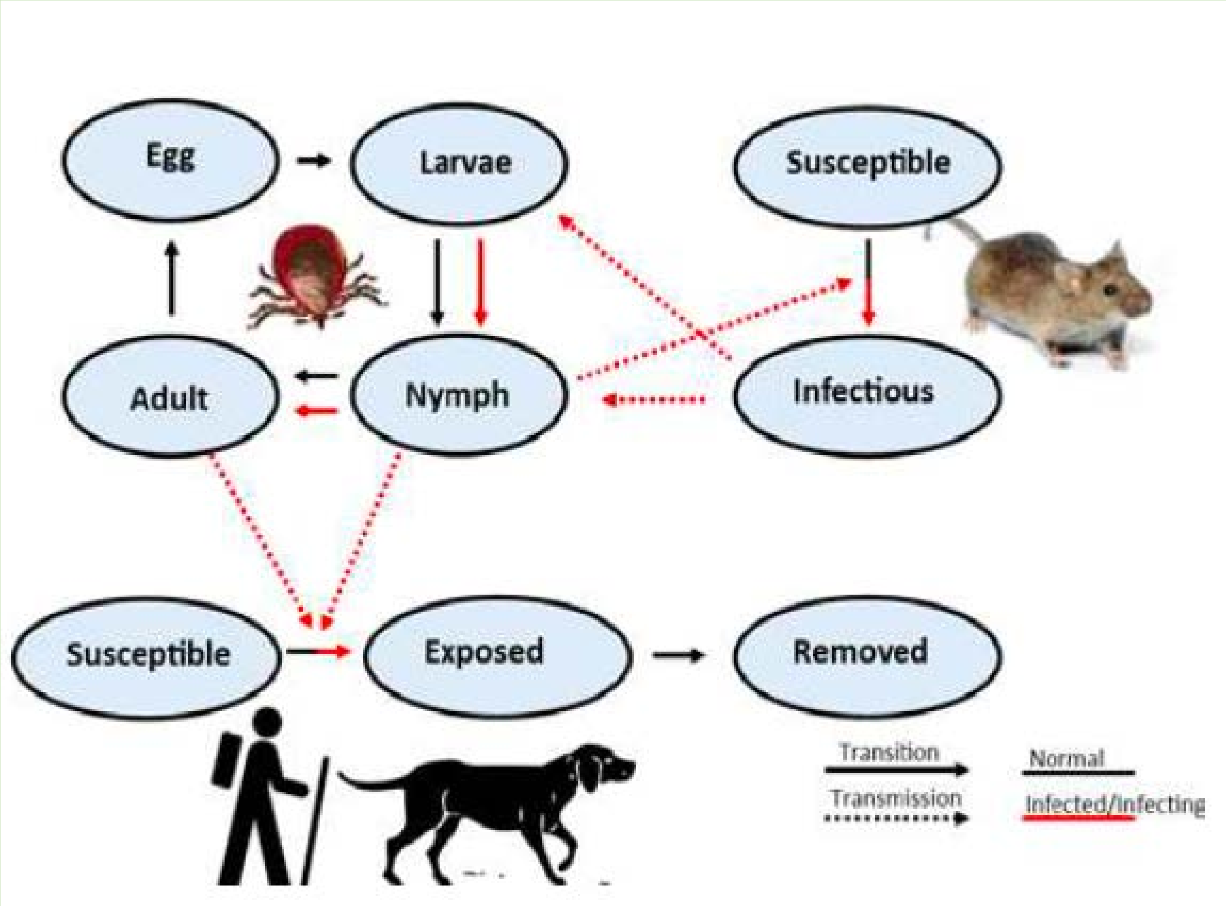
Ixodes scapularis

Taken by Grant Brown, PhD

Lyme Disease

- Most common disease spread by ticks in the Northern Hemisphere
  - Estimated to affect 300,000 people a year in the United States and 6,500 people a year in Europe
- Symptoms:
  - Redness on skin around site of tick bite
  - Rash, fever, headache, tiredness
  - If left untreated, loss of ability to move sides of face, joint pains, swelling

Lyme Disease Transmission



Graph provided by Peterson Lab

- Ixodes ticks transmit Lyme disease to dogs and humans
- Ticks have a 2-year life cycle with 4 life stages: egg, larva, nymph, and adult
- Ticks feed once before molting to next stage
- Black lines indicate stage transitions, red lines indicate disease transmission/infection

Project Goals

- Explore spatial distribution of Lyme cases
- Build models of Lyme incidence over time considering the effect of weather events, including precipitation, evaporation, and temperature extremes
- Evaluate forecasting ability of Lyme models

Data

- de-identified fully-insured administrative claims data from a private health insurance company between 2003-2016
  - The date the patient was first serviced was used as the diagnosis date
  - We were also interested in the diagnosis codes to determine cases of Lyme disease
  - Additionally, zip and county were used for mapping purposes

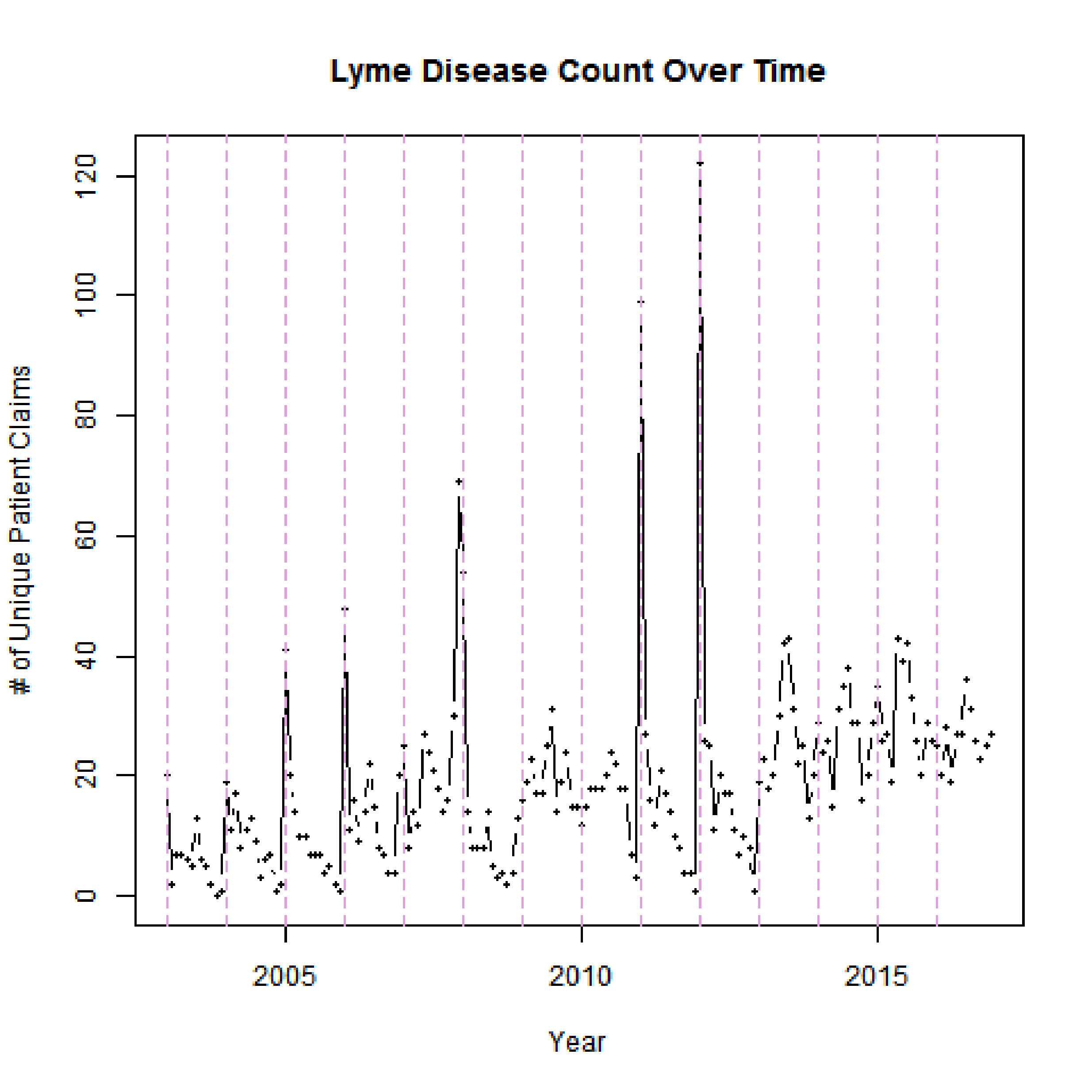
Data Continued

- Data from Global Historical Climatology Network-Daily Database
  - Taken from weather stations distributed across Iowa
  - Aggregated by month
  - We focused on each month's averages of:
    - Daily Minimum Evaporation
    - Daily Minimum Temperature
    - Daily Mean Precipitation

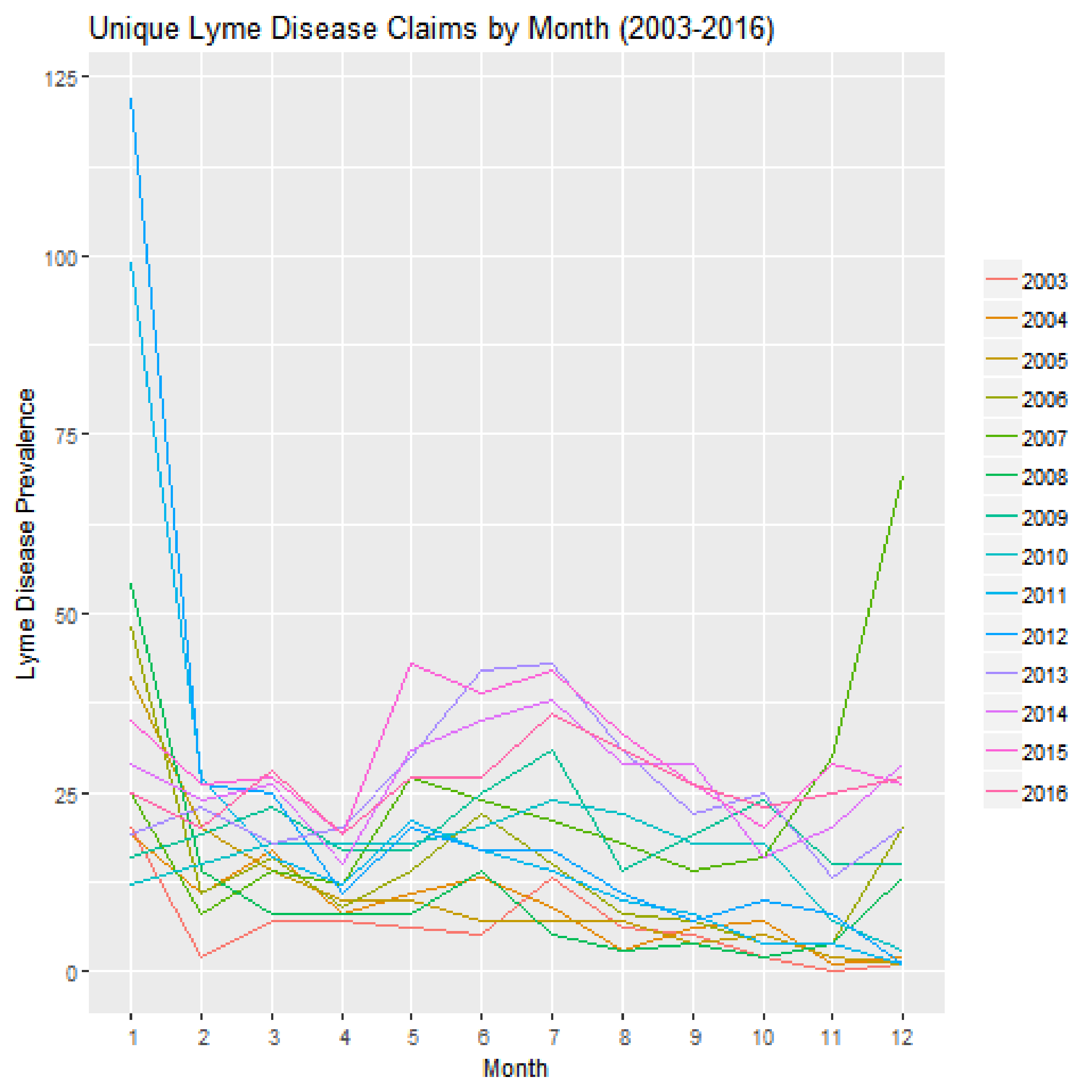
Data Processing

- The weather data was largely used as obtained
  - Evaporation missing values were interpolated using natural spline regression with df = 30 for model fitting
- We aggregated Health Insurance Data in several steps:
  - First, we read in one-year file at a time
  - Then, determined Lyme disease cases using claim codes
  - Grouped by individual patients to remove duplicate claims
  - Then grouped by county, year, and month and counted the number of claims
  - Finally, we aggregated the data

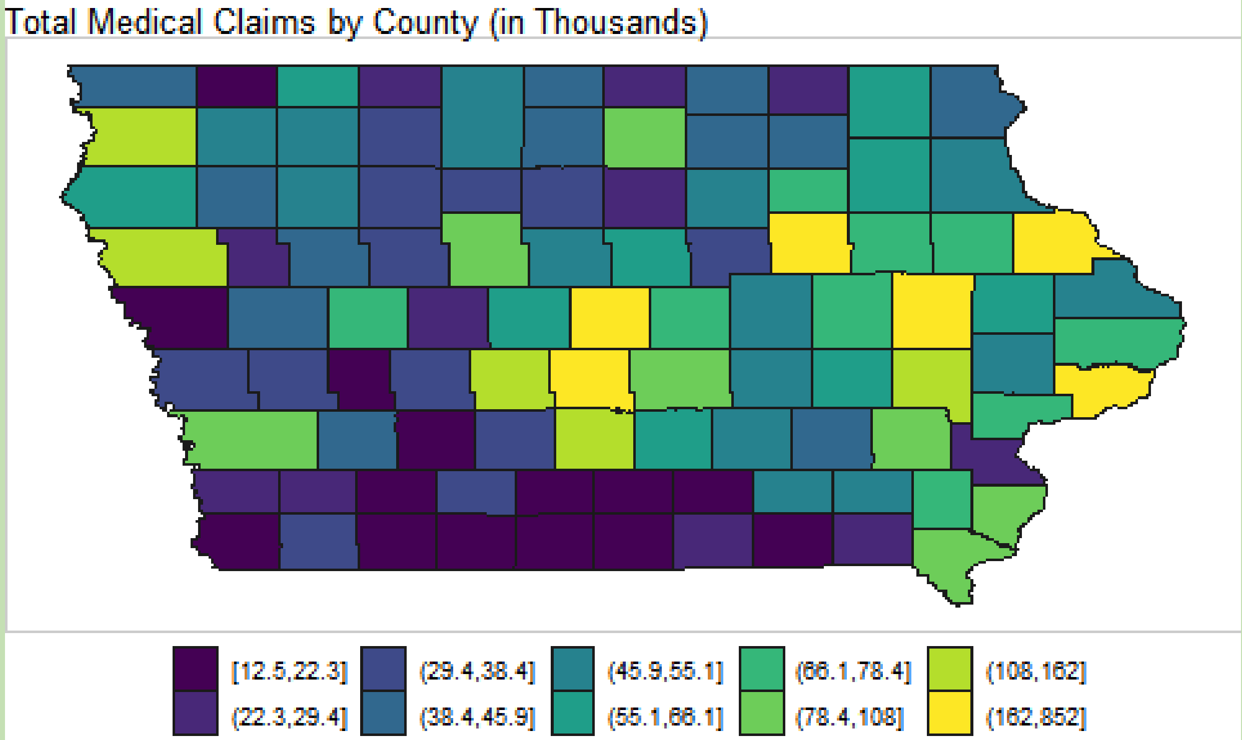
Time Series Visualization



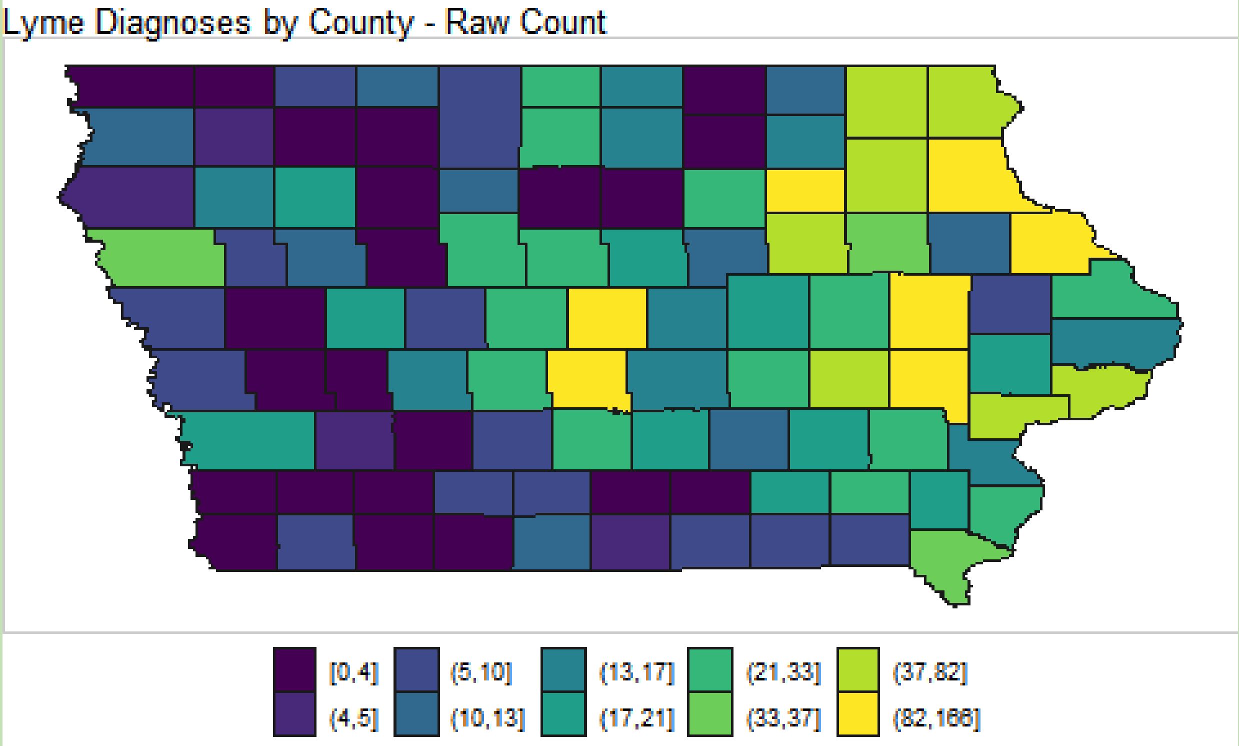
Times Series by Year



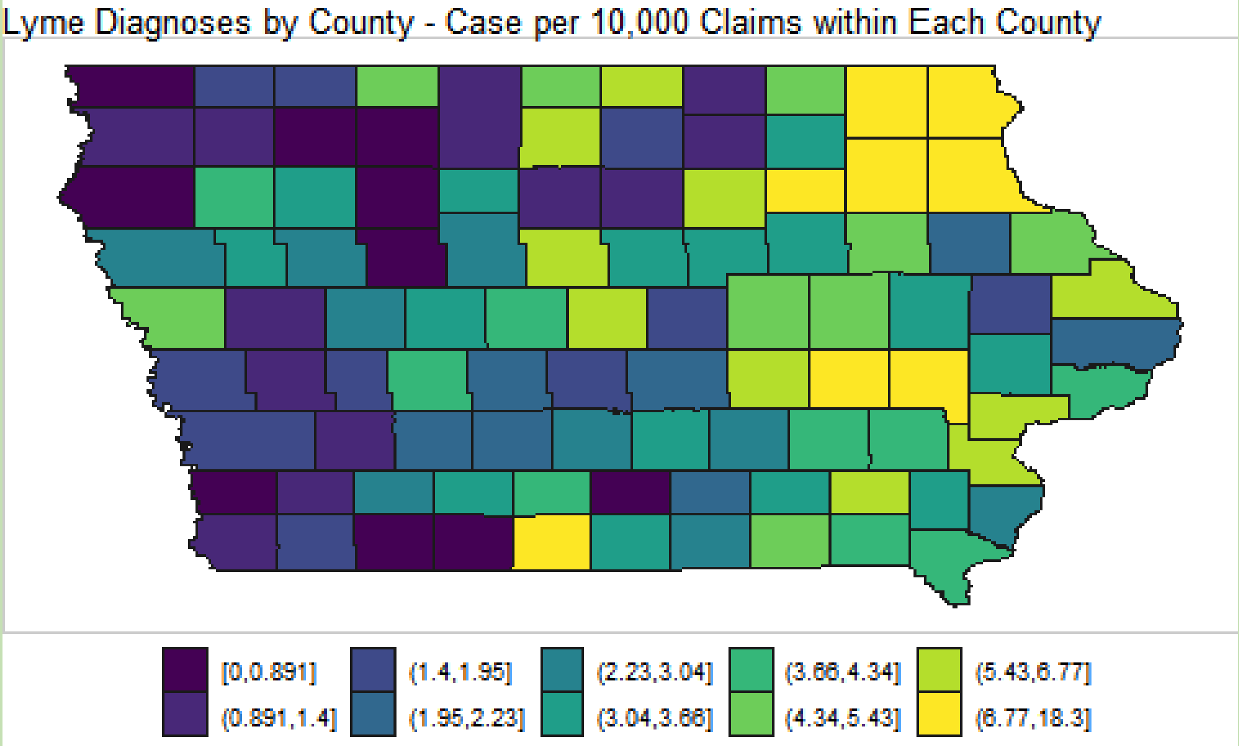
Iowa Insurance Claims by County



Iowa Lyme Diagnoses by County



Iowa Lyme Diagnoses by County: Proportion



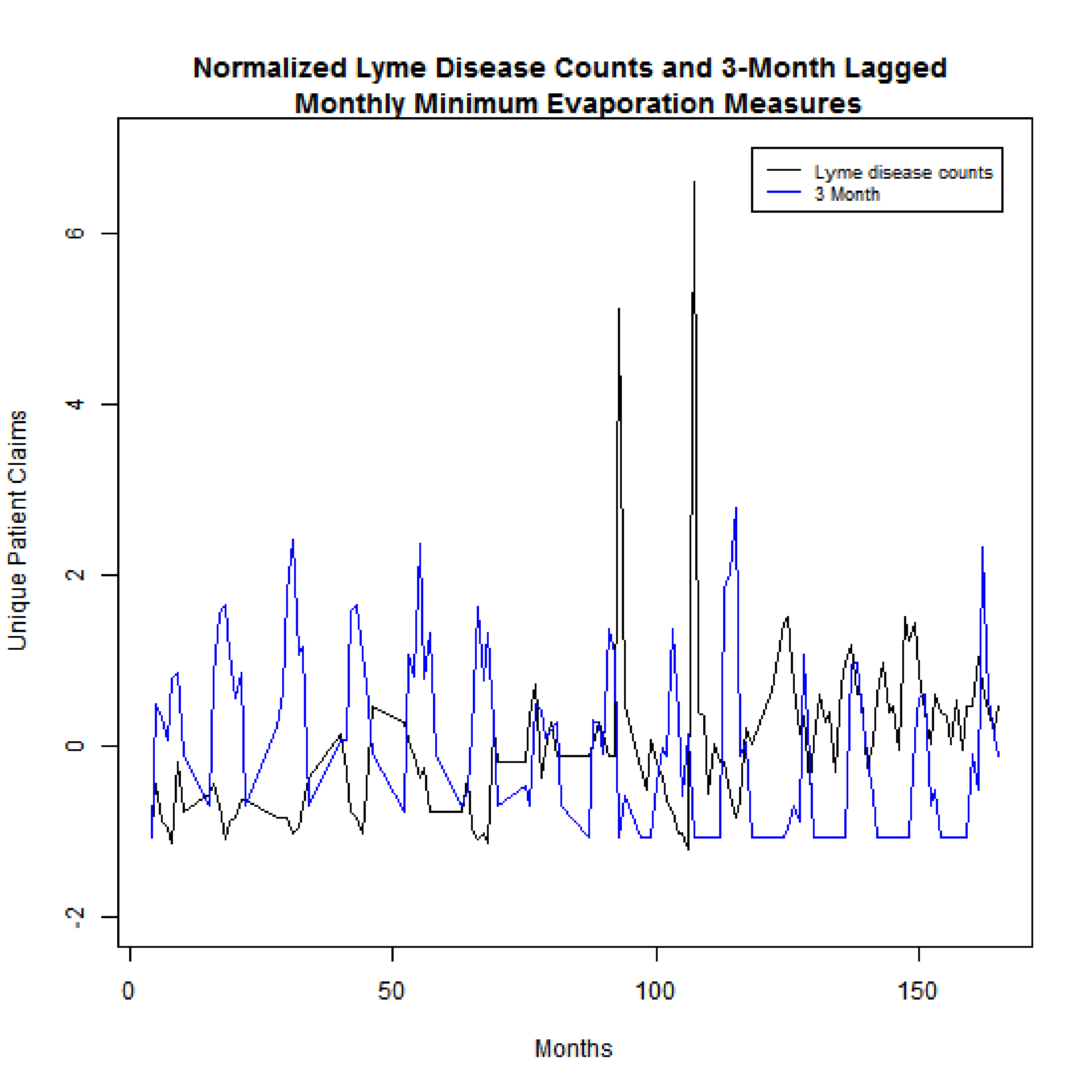
Lyme Disease Maps by Year

- Population adjusted Lyme disease counts (per 10,000 cases) https://www.youtube.com/watch?v=AUo-X8AdshM

Covariate Exploration

- We pulled 12 elements from the weather data, each of which had 4 variables, and we wanted to find the lagged correlations (including 0) up to 24 months.
- This would be 12*4*25 or 1200 correlations to explore
- We built functions to automatically standardize and lag each measure
- Then we looped over all possible lags to find maximum correlations with post-lagged data missing less than 30% of time points
- However, there are limitations of using the covariates

Time Series and Evaporation



Model Development

- The model needed to capture overall trend, seasonal trends, and impacts of climatic factors
- We faced two options and explored both initially: a Generalized Linear Model (GLM) or an Auto Regressive Moving Average (ARMA) model.
- GLM - Natural splines and trigonometric functions to model the time aspect with linear climactic covariates

$$Y_i \sim Poisson(\lambda)$$

$$\ln(\lambda_i) = \beta_0 + \beta_1 ns(t_i, df = 10) + \beta_2 sin(t_i) + \beta_3 cos(t_i) + \beta_4 sin(t_i) * \cos(t_i) + \beta_5 evp_i + \varepsilon_i$$

Model Development

- TS - Forecasting of future observations based on lagged observations and a moving average of residual errors from lagged observations
- The GLM has the approach of modeling the shape of the curve to account for seasonality with added weather covariates
- This model also assumes that each time point is independent
- On the other hand the time series model is dependent on previous time points and errors as well as exogenous variables allowing for greater flexibility
- The nature of the problem lends itself more naturally to time series

Auto-regressive moving average (ARMA) Model

- ARMA models can describe stochastic processes using two polynomials – one for auto-regression (AR) and one for the moving average (MA)
- Given a time series X_t, we can write an ARMA model as

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Where ε_t are error terms, c is a constant, and φ and θ are model parameters

NARMAX Model

- The previous model fails to capture seasonal trends
- We had two options:
  - Making the model dependent on lags 12 months prior
  - Adding exogenous variables that account for the seasonal trend of Lyme disease
- We chose the latter and implemented 0, 3, 6, 9, and 12-month lags for precipitation, minimum temperature, and evaporation
- The model was implemented using STAN and Bayesian methods

NARMAX Model Cont.

- Nonlinear auto-regressive moving average with exogenous inputs
- Recall from GLM,

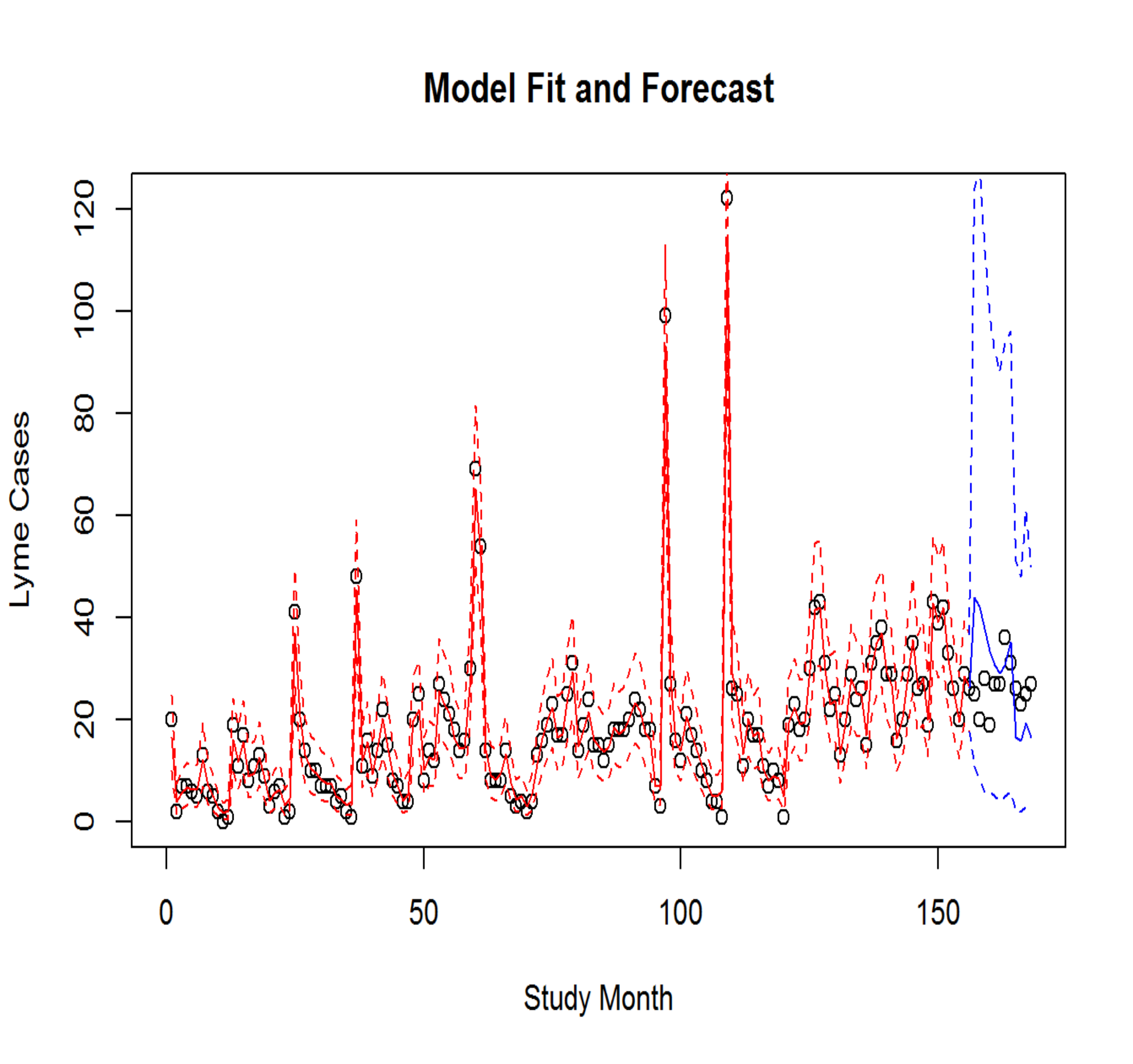
$Y_i \sim Pois(\lambda)$. Now, we have

$$\log(\lambda_i) = c + \varepsilon_i + \sum_{\mathrm{j}=i-1}^{i-k} \varphi_i \log(\lambda_i) + \sum_{\mathrm{j}=i-1}^{i-k} \theta_i \varepsilon_j + \underbrace{\sum_{j=i-1}^{i-k} \sum_{l=1}^{3} X_j^{(l)} \beta_l}_{\text{“past weather”}}$$

Model Assumptions

- Assumes we can predict weather trends accurately 12-months out
- Assumes NARMAX (1,1) sufficient
- Assumes Poisson sufficient; mean equals variance
- Assumes no missing covariates
- Assumes latent ARMA variance is constant

NARMAX Model



Results

Precipitation

	Mean	Median	2.5%	97.5%	Prob <> 0
Current	-0.0006	-0.0006	-0.0086	0.0075	0.5604
3 month	-0.0039	-0.0040	-0.0123	0.0045	0.8219
6 month	-0.0056	-0.0056	-0.0138	0.0026	0.9114
9 month	0.0028	0.0028	-0.0050	0.0105	0.7600
12 month	-0.0016	-0.0016	-0.0091	0.0060	0.6606

Temperature Minimum

	Mean	Median	2.5%	97.5%	Prob <> 0
Current	0.0029	0.0029	0.0004	0.0054	0.9878
3 month	-0.0002	-0.0002	-0.0026	0.0022	0.5757
6 month	-0.0008	-0.0008	-0.0034	0.0019	0.7096
9 month	0.0000	0.0000	-0.0025	0.0026	0.5030
12 month	-0.0042	-0.0042	-0.0069	-0.0015	0.9989

Evaporation

	Mean	Median	2.5%	97.5%	Prob <> 0
Current	0.0009	0.0010	-0.0116	0.0134	0.5611
3 month	-0.0145	-0.0144	-0.0265	-0.0025	0.9905
6 month	0.0084	0.0084	-0.0041	0.0211	0.9058
9 month	-0.0003	-0.0003	-0.0118	0.0114	0.5237
12 month	0.0089	0.0089	-0.0038	0.0213	0.9175

- Most significant factors in modeling/predicting Lyme cases are the 6-month average for precipitation, the average minimums of a 0 and 12-month lag for minimum temperature, and a 3, 6, and 12-month lag for evaporation.
- We also had an alpha term that captured the linear trend. It had a .999 probability of being greater than zero with a 95% credible interval of [.00317, .01221]

Discussion

- The large prediction interval is likely due to the large deviations often seen in January which are unaccounted for in the exogenous variables and thus have nowhere to go but the error term
- This could be due to the underlying process of capturing claims data
- Further exploration to account for variability would help reduce the credible interval

Conclusion

- Further study needs to be done on working with claims data
- Additional work needed to find best predictors of Lyme
- Future models can tie in spatiotemporal work with weather data using different techniques

Acknowledgements

Grant Brown, PhD

Gideon Zamba, PhD

Terry Kirk, Graduate Admin

Melissa Jay

Biyue Dai

Helín Hernandez

National Heart, Lung, and Blood Institute (NHLBI), grant #HL147231.

References

- https://en.wikipedia.org/wiki/Lyme_Disease
- Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily Database. Journal of Atmospheric and Oceanic Technology, 29, 897-910, doi:10.1175/JTECH-D-11-00103.1.
- Menne, M.J., I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B.E.Gleason, and T.G. Houston, 2012: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.
- NOAA National Climatic Data Center. http://doi.org/10.7289/V5D21VHZ [July 1, 2019].