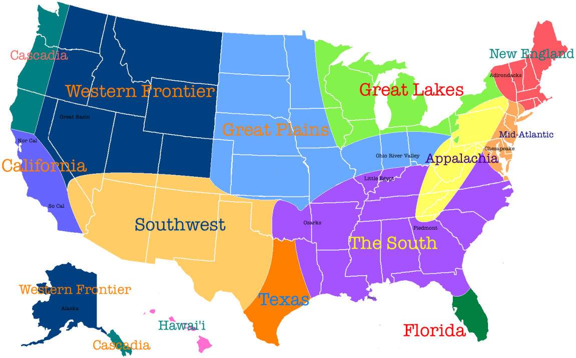
Using Bayesian Spatiotemporal Modeling to Understand Mortality Rates in the United States

Brian Hartman
Joint Work with Robert Richardson, Chris Groendyke, Zoe Gibbs, McKay Christensen, Michael Shull, and Jared Cummings

Brigham Young University

November 2022

Introduction



- Life varies by region in both measurable and immeasurable ways
- We want to improve existing mortality models by accounting for spatiotemporal trends

The Data

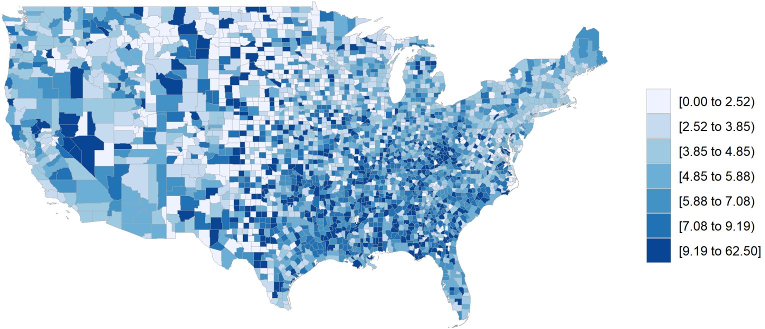
- Characteristics of every death in the United States from 2000-2017
- Used data from the US decennial census and the American Community Surveys from intercensus years for population exposures
- Census estimates are binned by gender and 18 five-year age groups from 0 to 85 years old

The Data

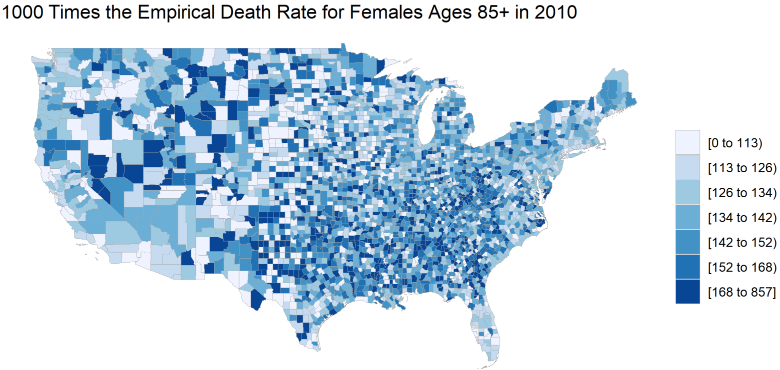
- Used data from the contiguous United States to maintain spatial relationships
- Combined counties with extremely low populations and to account for boundary changes, leading to 3,092 total counties for study
  - 22,590,587 Female Deaths
  - 22,446,212 Male Deaths

Exploratory Data Analysis

1000 Times the Empirical Death Rate for Females Ages 55-59 in Year 2010



Exploratory Data Analysis



Exploratory Data Analysis

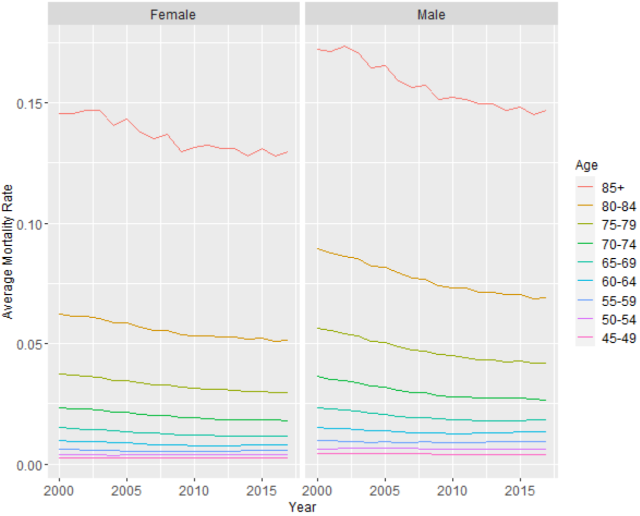


Figure 1: Average mortality rate by gender, age group, and year (for older ages)

Exploratory Data Analysis

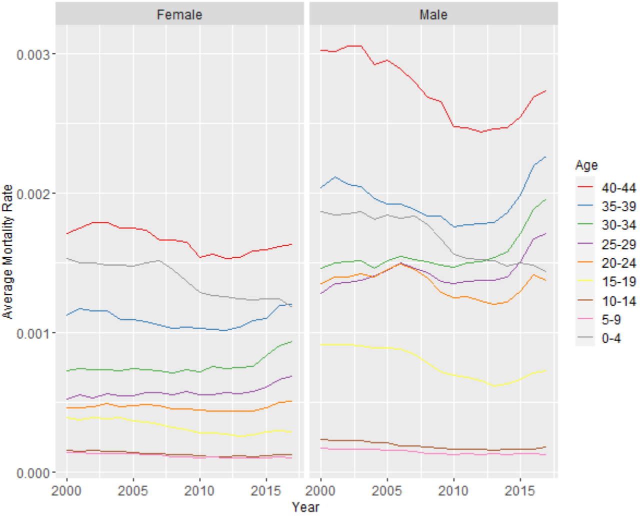
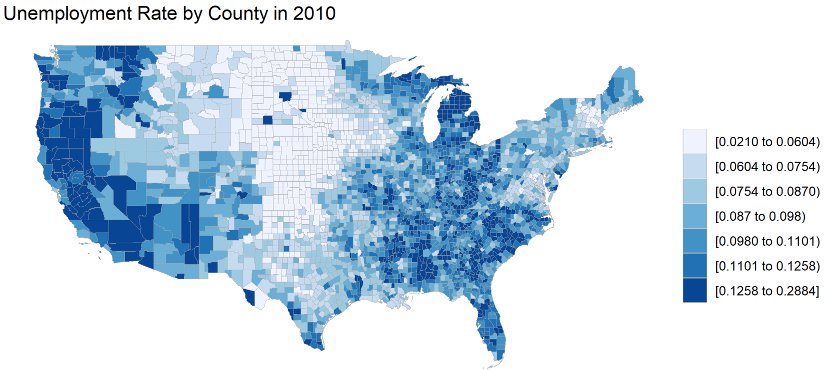


Figure 2: Average mortality rate by gender, age group, and year (for younger ages)

The Data

- Considered several demographic and economic statistics as explanatory variables
- We preferred complete data for every county and year
- Used unemployment rate estimates as provided by the US Bureau of Labor Statistics
- We hope to obtain additional explanatory variables in the future

Exploratory Data Analysis



Exploratory Data Analysis

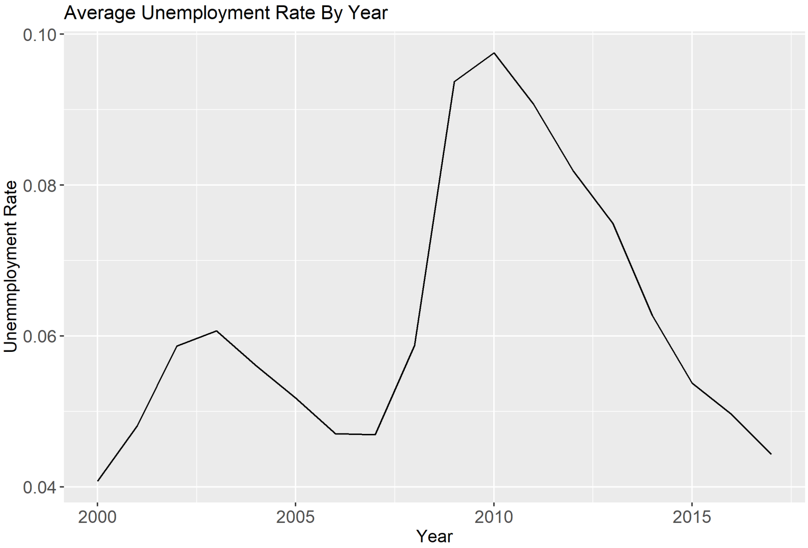


Figure 3: Average unemployment rate, weighted by county population in 2000-2017

Bayesian CAR Models

- Model with a prior such that the every space/time point only depends on direct neighbors and adjacent time points
- Special case of a Markov random field
- Often used to analyze areal data

Bayesian Binomial Hierarchical Model

$$y_{kt} \sim \text{Binomial}(n_{kt}, \theta_{kt}) \quad (1)$$

$$\log\left(\theta_{kt}/(1-\theta_{kt})\right) = \mathrm{x}'_{kt}\boldsymbol{\beta} + \psi_{kt} \quad (2)$$

- n_{kt} is the total population and θ_{kt} is the probability of death.
- $\mathrm{x}'_{\mathrm{kt}}$ represents the covariate information for the k^{th} location at time t
- $\boldsymbol{\beta}$ is a vector of coefficients
- ψ_{kt} collects all the spatial and temporal random effects that create the spatio-temporal dependence

Conditional Autoregressive Priors

$$y_i | y_{(i)} \sim N\left(\rho \sum_{j=1}^{n} \frac{1}{N_i} W_{ij} y_j, \tau^2\right) \tag{3}$$

$$\mathrm{Y} \sim N\left(0, \tau^2 (D - \rho W)^{-1}\right) \tag{4}$$

- $y_{(i)}$ represents all locations excluding location i and N_i is the total number of neighbors for location i
- ρ represents the degree of dependence between neighbors
- W is a location matrix such that if W_{ij} equals 1 if and only if i and j are direct neighbors and 0 otherwise
- $D = \mathrm{diag}(N_1, ..., N_n)$ is a diagonal matrix collecting the number of neighbors for each location.

CAR Linear Model

$$\psi_{kt} = \phi_k + (\alpha + \delta_k)\frac{t - \bar{t}}{T} \tag{5}$$

- Assumes a linear trend in the random effect over time with a slope equal to $\frac{\alpha+\delta_k}{T}$
- ϕ_k and δ_k are given CAR priors according to Equation 4

CAR Linear Model

- The other parameters are given standard priors.

$$
\begin{aligned}
\boldsymbol{\phi} &\sim N\left(0, \tau_s^2 (D - \rho_s W)^{-1}\right) && (6)\\
\boldsymbol{\delta} &\sim N\left(0, \tau_t^2 (D - \rho_t W)^{-1}\right) && (7)\\
\tau_s^2, \tau_t^2 &\sim \mathrm{IG}(1, 0.01) && (8)\\
\rho_s, \rho_t &\sim \mathrm{Uniform}(0, 1) && (9)\\
\beta_0, \beta_1 &\sim N(0, 1) && (10)
\end{aligned}
$$

Computational Technique

- We used the **CARBayesST** package in R to perform MCMC sampling from the posterior
- We ran separate models for each age group and gender combination for a total of 36 models
- $\boldsymbol{\beta}$, ϕ_k, and δ_k are updated using the Metropolis algorithm with a normal proposal distribution
- τ_S^2, τ_T^2 are updated using conjugacy principles and Gibbs sampling
- ρ_S, ρ_T are updated using the Metropolis-Hastings algorithm with a truncated normal proposal distribution
- Trace plots and Geweke diagnostics were used to assess convergence

Results

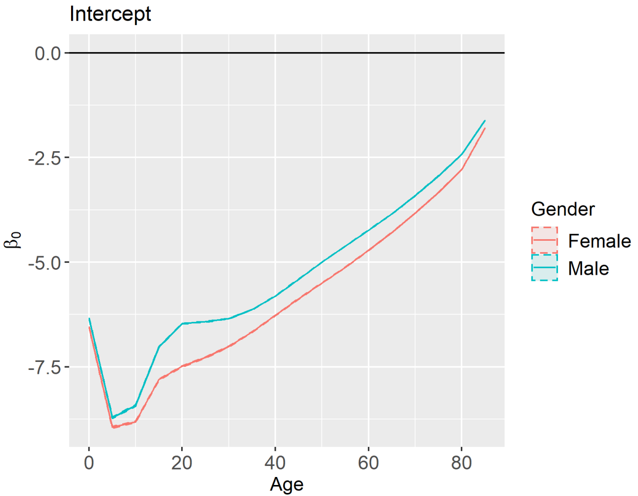


Figure 4: Estimates and 95% credible intervals for β_0 (intercept) for each age group and gender combination.

Results

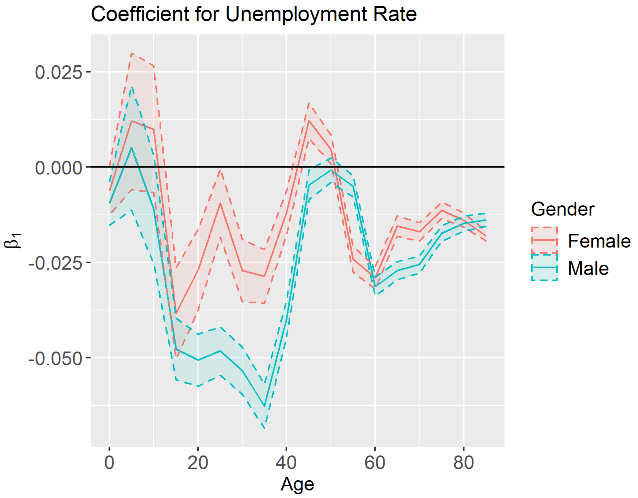


Figure 5: Estimates and 95% credible intervals for β_1 (coefficient for unemployment rate) for each age group and gender combination.

Results

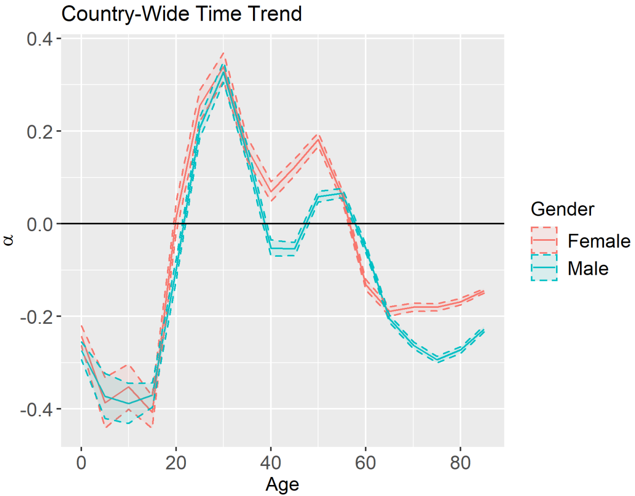


Figure 6: Estimates and 95% credible intervals for α (country-wide time trend) for each age group and gender combination.

Results

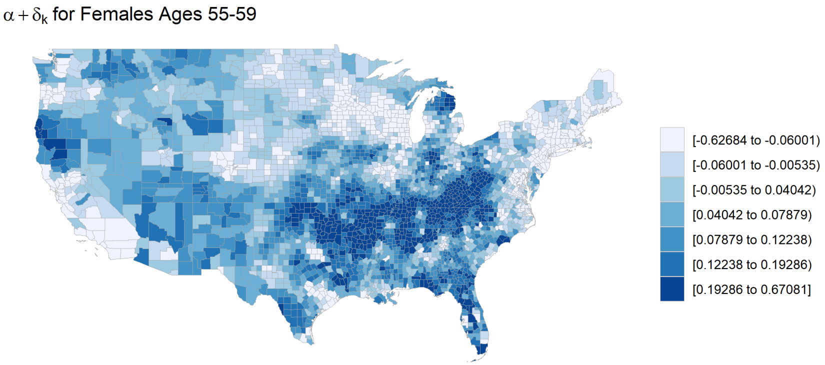
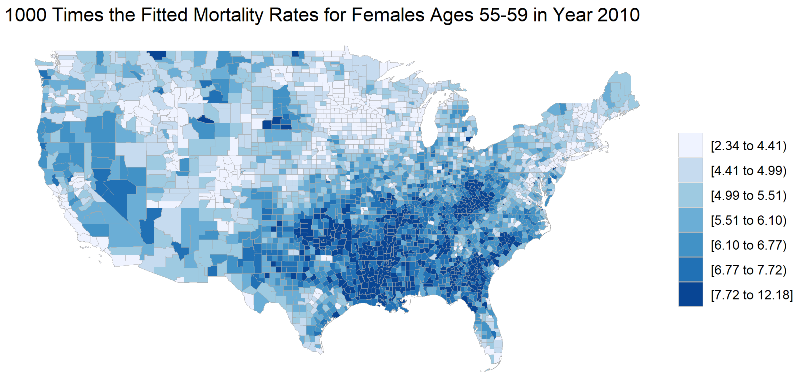


Figure 7: Estimates for $\alpha + \delta_k$ (total time trend) by county for Females ages 55-59

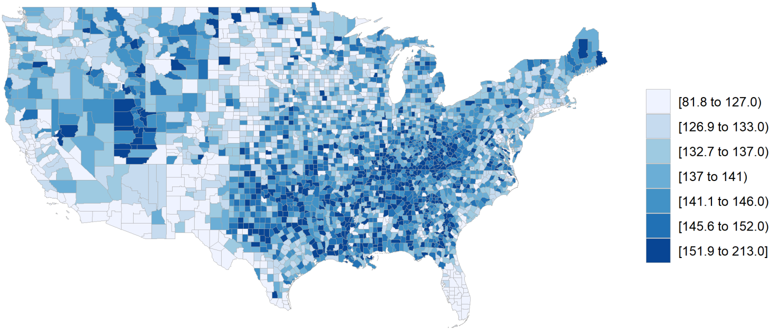
Results



1000 Times the Fitted Mortality Rates for Females Ages 55-59 in Year 2010
[2.34 to 4.41)
[4.41 to 4.99)
[4.99 to 5.51)
[5.51 to 6.10)
[6.10 to 6.77)
[6.77 to 7.72)
[7.72 to 12.18]

Results

1000 Times the Fitted Mortality Rates for Females Ages 85+ in Year 2010



Model 2

To further understand the spatial relationships of mortality rates, we built a slightly more complicated model to examine state-specific covariate effects.
We again start with a binomial distribution as the top level of the hierarchy.

$$y_{akt}|\pi_{akt} \sim \text{Binomial}(n_{akt}, \pi_{akt})$$

for each age group a in county k during year t.

Model 2

We can then relate π_{akt} to the desired effects using the logit link function:

$$\ln\left(\frac{\pi_{akt}}{1-\pi_{akt}}\right) = \beta_0 + \sum_{i=1}^{3} F_i(\mathrm{x}_k) + \sum_{i=1}^{3} G_{is}(\mathrm{x}_{kt}) + \phi_k + \delta_t + \psi_a + \gamma_{akt}$$

where

- β_0 is the intercept.
- Each F_i and G_{is} is a nonlinear covariate effect modeled by a Gaussian process with Matérn correlation function.
- $\phi_k = u_k + v_k$, where u_k is the structured (CAR) and v_k is the unstructured spatial effect.
- δ_t is the temporal effect, ψ_a is the age group effect, and γ_{akt} is the error term.

Model 2 Computation

We use integrated nested Laplace approximations (INLA) to fit our model making it computationally feasible.

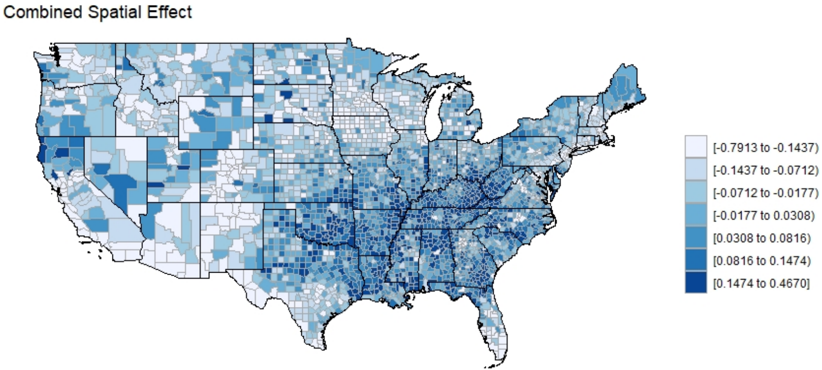
Model Selection

We compared three different models using DIC to make sure the added complication was worthwhile.

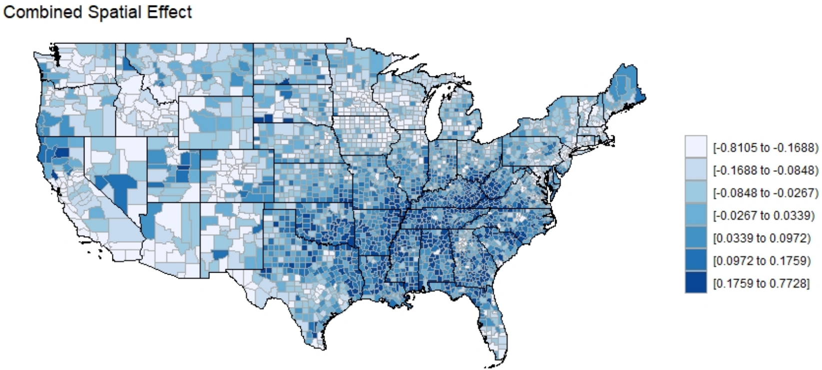
Model	DIC (Female)	DIC (Male)
Full Model	3,817,853	4,266,276
Only Countrywide	3,818,372	4,266,666
No Covariates	3,819,075	4,266,790

Table 1: Deviance Information Criterion (DIC) for the three different model versions that were fit to both the male and female data.

Model 2 Overall Female Spatial Effect



Model 2 Overall Male Spatial Effect



Combined Time Effect

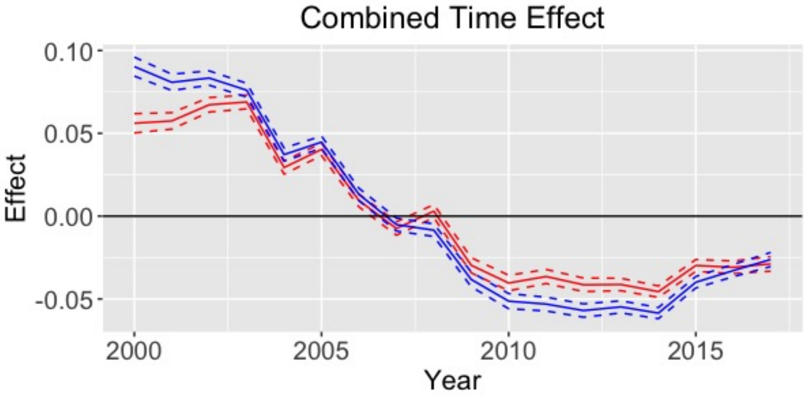


Figure 8: Posterior mean and 95% credible interval of the temporal effects (δ_t). Male values are in blue and female values are in red.

Combined Age Effect

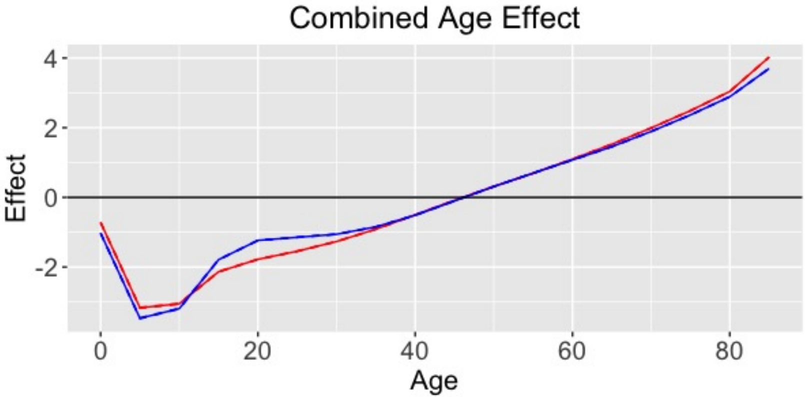
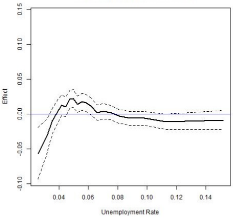
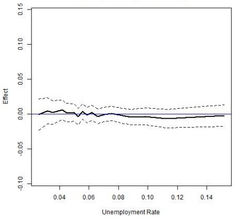


Figure 9: Posterior mean and 95% credible interval of the age group effects (ψ_t). Male values are in blue and female values are in red.

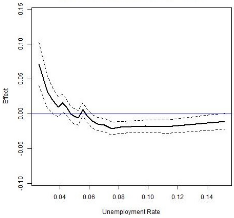
State-specific Unemployment Effect



(a) California



(b) Colorado



(c) North Carolina

Figure 10: Posterior mean and 95% credible interval of the unemployment effects ($G_{1s}(x_{kt})$) for selected states for the model fit to the female data.

Conclusions and Future Work

- Incorporating spatial correlation into mortality modeling can help us better understand mortality rates
- The spatial dependence parameters helps us draw on information from neighboring counties
- In the future we would like to incorporate additional covariates
- We also want to compare other models and techniques

Using Bayesian Spatiotemporal Modeling to Understand Mortality Rates in the United States

Brian Hartman
Joint Work with Robert Richardson, Chris Groendyke, Zoe Gibbs, McKay Christensen, Michael Shull, and Jared Cummings

Brigham Young University

November 2022