Bayesian Spatio-temporal analysis of mortality differentials in the US using the INLA approach

Corey S. Sparks, Ph.D.

Univerity of Texas at San Antonio - Department of Demography

May 15, 2018

Presentation Structure

- Introduction to Mortality Disparities in the US
- The role of residential segregation
- Research Questions
- Data
- Methods - The INLA approach to Bayesian analysis
- Results & visualizations
- Wrap up

Introduction

- Black-White disparities in mortality rates persist
- Most research focuses on individual level factors
  - SES, Health behaviors
- More recent work is multilevel
  - Context of health, neighborhoood conditions
- Role of residential segregation on aggregate mortality rates still poorly understood

Segregation & Mortality

- Williams and Collins (2001) offer one of the first conceptual pieces to link segregation to poor health.
- Segregation spatially and socially patterns:
  - Poverty
  - Economic and educational opportunities
  - Social order or disorder
  - Access to resources
- Segregation could lead to better health outcomes (political representation, social support, cohesion)

Research Questions

- Does the effect of segregation produce the same disparity in black and white mortality rates over time?
- Do counties with persistently high segregation show the same mortality disadvantage for both black and white mortality rates?
- Does segregation have any protective advantage on county-level mortality rates?
  - For black mortality specifically

Data

- NCHS Compressed Mortality File
  - County - level counts of deaths by year, age, sex, race/ethnicity and cause of death
  - 1980 to 2010
  - Age, sex and race *(white & black)* specific rates for all US counties
  - In total: 35748276 deaths in the data
  - Standardized to 2000 Standard US population age structure
  - Rates stratified by race and sex for each county by year
  - n = 2 sexes * 2 races * 3106 counties * 31 years = 385144 observations
  - *Analytic* n = 315,808 nonzero rates

Data - Access

- You can basically get these data from the CDC Wonder website
- Supresses counts where the number of deaths is less than 10
- Rates are labeled as "**unreliable**" when the rate is calculated with a numerator of 20 or less
  - Big problem for small population counties
  - Still a problem for large population counties!
- Restricted use data allows access to **ALL** data

Data - Example

Bexar County, TX 1980 - 1982

County	Year	Race-Sex	Rate
48029	1980	White Female	7.920
48029	1980	Black Female	9.960
48029	1980	White Male	13.508
48029	1980	Black Male	17.179
48029	1981	White Female	8.216
48029	1981	Black Female	9.822
48029	1981	White Male	12.870
48029	1981	Black Male	15.442
48029	1982	White Female	7.592
48029	1982	Black Female	10.072
48029	1982	White Male	12.894
48029	1982	Black Male	16.663

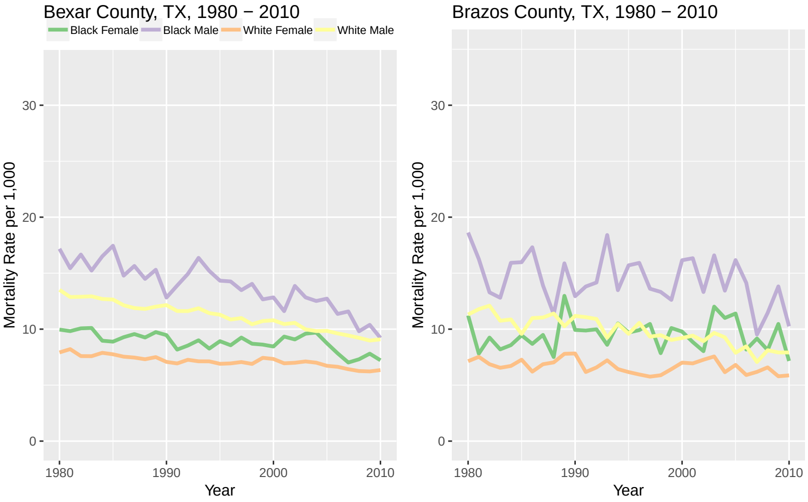
Data - Example

Brazos County, TX 1980 - 1982

County	Year	Race-Sex	Rate
48041	1980	White Female	7.138
48041	1980	Black Female	11.219
48041	1980	White Male	11.308
48041	1980	Black Male	18.630
48041	1981	White Female	7.527
48041	1981	Black Female	7.812
48041	1981	White Male	11.788
48041	1981	Black Male	16.263
48041	1982	White Female	6.867
48041	1982	Black Female	9.246
48041	1982	White Male	12.096
48041	1982	Black Male	13.284

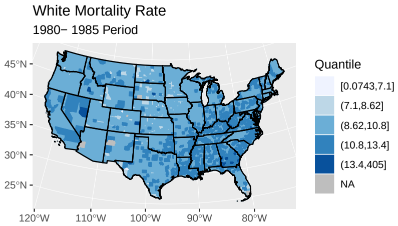
Data - Example

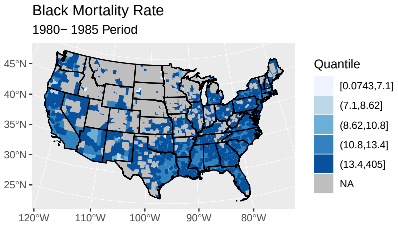
County specific temporal trends 1980 - 2010



Data - Example of Geographic Variation

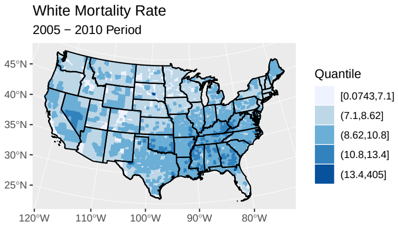
Spatial Distribution of White & Black Mortality in the US: 1980-1985 Period

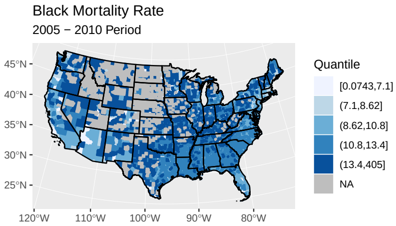




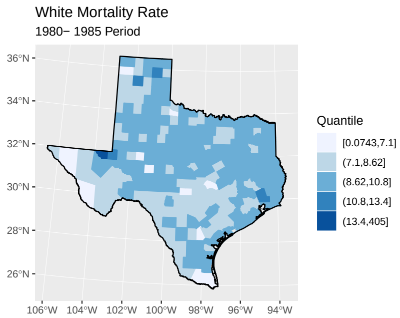
Data - Example of Geographic Variation

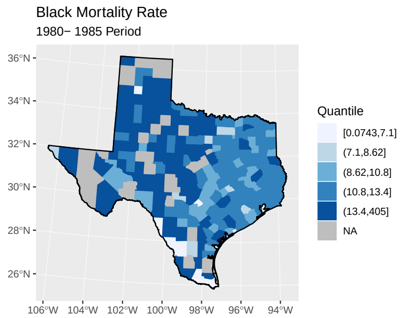
Spatial Distribution of White & Black Mortality in the US: 2005-2010 Period





State Examples





Methods - Hierarchical Model

- I specify a Bayesian Hierarchical model for the age-standardized mortality rate
- $$Y_{ij} \sim N(\mu_{ij}, \tau_y)$$
- $$\mu_{ij} = \beta_0 + x'\beta + \gamma_j * black_i + u_j +$$
- $$\gamma_1 * \text{time} + \gamma_2 * (\text{time} * black_i) + \gamma_3 * (\text{time} * seg_i)$$
- $$\gamma_j \sim \text{CAR}(\bar{\gamma}_j, \tau_\gamma / n_j)$$
- $$u_j \sim \text{CAR}(\bar{u}_j, \tau_u / n_j)$$
- Vague Gamma priors for all the τ's
- Vague Normal priors for all the fixed effect β's and γ's

Methods - Bayesian analysis

- This type of model is commonly used in epidemiology and public health
- Various types of data likelihoods may be used
- Need to get at:

*

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

- Traditionally, we would get $p(\theta|y)$ by:
  - either figuring out what the full conditionals for all our model parameters are (hard)
  - Use some form of MCMC to arrive at the posterior marginal distributions for our parameters (time consuming)

Methods - INLA approach

- Integrated Nested Laplace Approximation - Rue, Martino & Chopin (2009)
- One of several techniques that approximate the marginal and conditional posterior densities
  - Laplace, PQL, E-M, Variational Bayes
- Assumes all random effects in the model are latent, zero-mean Gaussian random field, x with some precision matrix
  - The precision matrix depends on a small set of hyperparameters
- Attempts to construct a joint Gaussian approximation for $p(x|\theta, y)$
  - where θ is a small subset of hyper-parameters

Methods - INLA approach

- Apply these approximations to arrive at:
- $\tilde{\pi}(x_i|y) = \int \tilde{\pi}(x_i|\theta, y)\tilde{\pi}(\theta|y)d\theta$
- $\tilde{\pi}(\theta_j|y) = \int \tilde{\pi}(\theta|y)d\theta_{-j}$
- where each $\tilde{\pi}(.|.)$ is an approximated conditional density of its parameters
- Approximations to $\pi(x_i|y)$ are computed by approximating both $\pi(\theta|y)$ and $\pi(x_i|\theta, y)$ using numerical integration to integrate out the nuisance parameters.
  - This is possible if the dimension of θ is small.
- Approximations to $\tilde{\pi}(\theta|y)$ are based on the Laplace appoximation of the marginal posterior density for $\pi(x, \theta|y)$
- Their approach relies on numerical integration of the posterior of the latent field, as opposed to a pure Gaussian approximation of it

INLA in R

```
library(INLA)
```

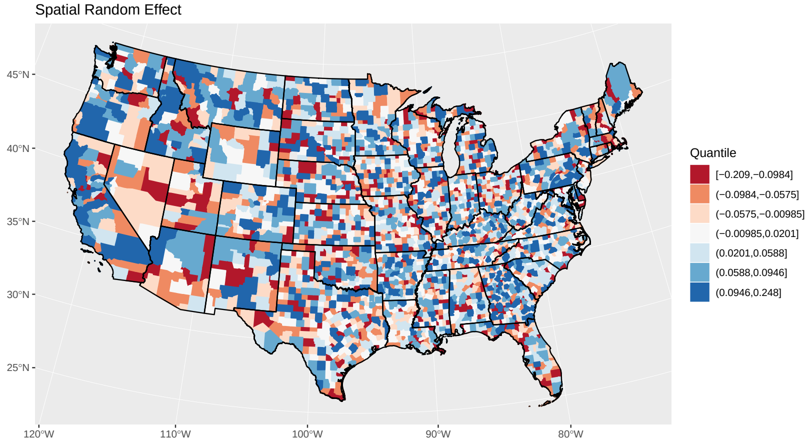
Unstructured Model

```
mod1<-std_rate~male+black+scale(lths)+pershigdis*year
+f(year,model="iid") +f(conum, model="iid")
```

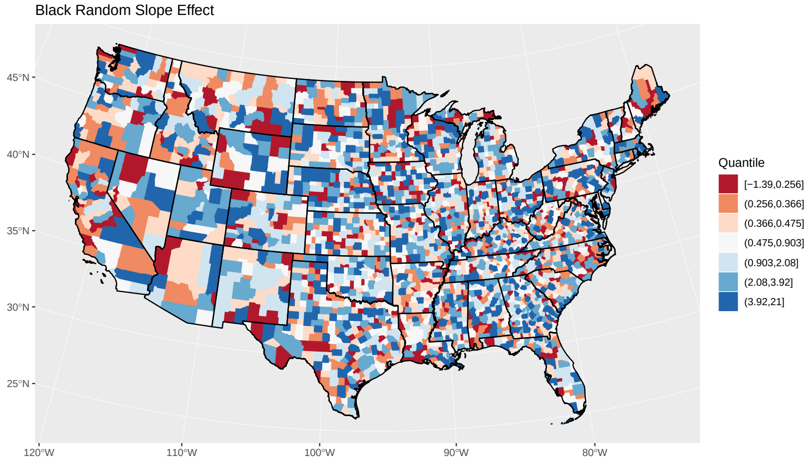
Spatially structured Model with Random Slope

```
mod2<-std_rate~male+black+scale(lths)+pershigdis*year
+f(conum, model="bym", graph="usagraph.gra") +f(year,
model="iid") +f(year, black,
model="besag",graph="usagraph.gra")
```

Spatial Model Results



```
p2
```



Black Random Slope Effect
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
Quantile
[−1.39,0.256]
(0.256,0.366]
(0.366,0.475]
(0.475,0.903]
(0.903,2.08]
(2.08,3.92]
(3.92,21]

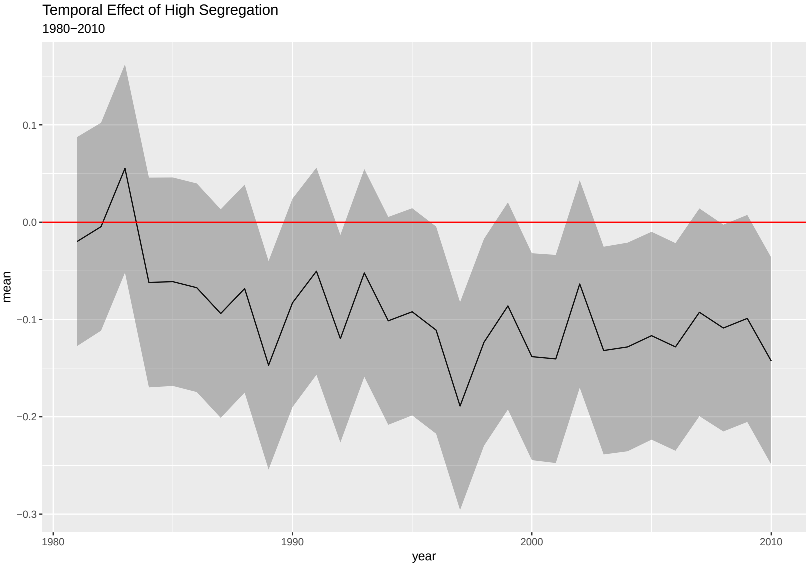
Model Results

- Fixed effects

```
##                    Post.Mean Lower_BCI Upper_BCI
## Male                   0.573     0.568     0.579
## Black                  1.761     1.708     1.814
## County_Low_Edu         0.027     0.021     0.034
## High_Segregation       0.109     0.032     0.187
```

```
##                Post.Mean Lower_BCI Upper_BCI
## Gaussian var  0.558547  0.561344  0.555658
## Spatial var   0.003852  0.004682  0.003213
## Black RS var 22.391685 23.566835 21.465436
```

Temporal effects of segregation



Discussion

- We see that, while there is a persistence of the gap in black-white mortality:
  - The mortality gap appears to be fairly consistent over time
  - In highly segregated areas, the mortality difference are decreasing
  - Suggests some evidence to support the Williams and Collins (2001) perspective
- INLA allows for rapid deployment of Bayesian statistical models with latent Gaussian random effects
  - Faster and *generally* as accurate as MCMC
  - Potentially an attractive solution for problems where large data/complex models may make MCMC less desireable

Thank you!

corey.sparks@utsa.edu

Coreysparks1

UTSA Demography