Assessing the Descriptive Epidemiology of Idiopathic Clubfoot in Iowa

Siri Neerchal

University of Maryland, College Park

Mentors: Dr. Jake Oleson & Dr. Paul Romitti

University of Iowa

July 19, 2018

Overview

What is Clubfoot?

- Also known as talipes equinovarus
- One or both feet are rotated inward
- Common musculoskeletal birth defect ($\sim$ 1 case per 1000 births)
- Causes: unknown

Prior Research

- Iowa Registry for Congenital and Inherited Disorders (IRCID) surveillance data
- Previous epidemiological work in Iowa: 1997-2005 (Kancherla et al.)
- Known associations:
  - Higher prevalence in males
  - Carter effect
  - Exposure to smoking in utero

Project Objectives

- Examine trends in clubfoot prevalence in Iowa
- Estimate associations with selected child and parental characteristics
- Identify geographic hotspots in the state

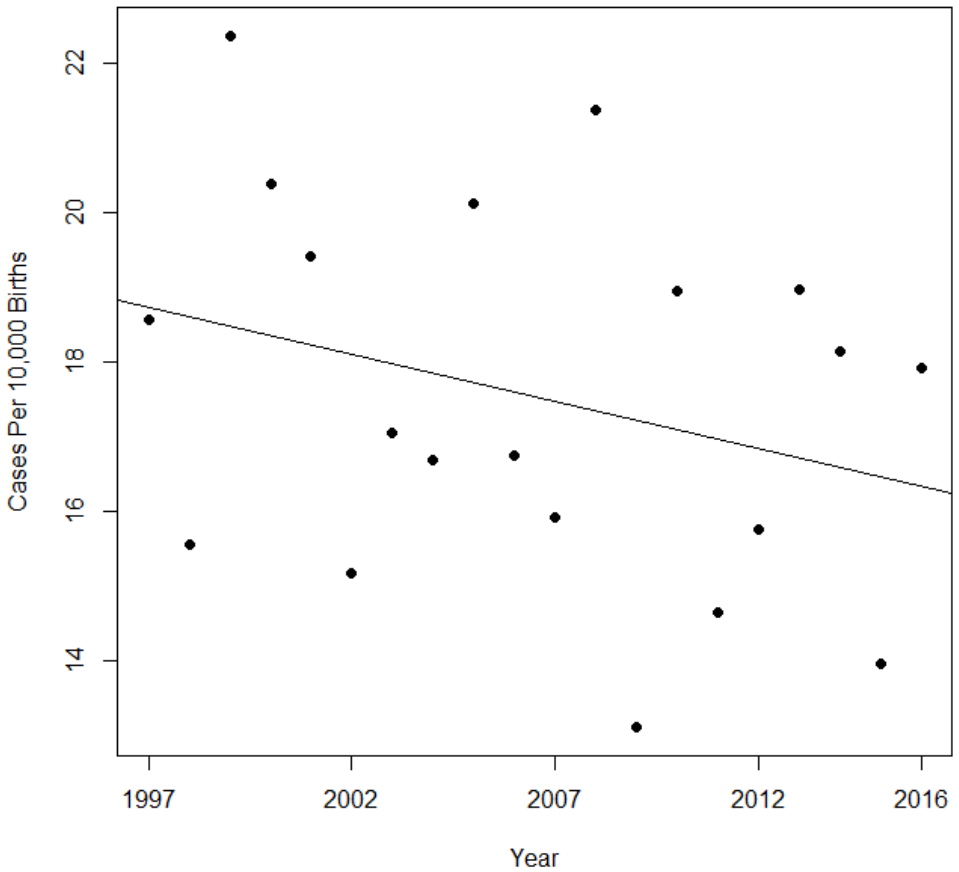
Clubfoot Cases in Iowa, 1997-2016

(Subsets Considered)

Total	Live	Idiopathic	Bilateral	Unilateral
1358	1194	783	387	396

All Live Births (1997-2016): 774,769

Prevalence over Time



$$\hat{Y} = 173.125 - 0.081X$$

p-values: 0.136, 0.159

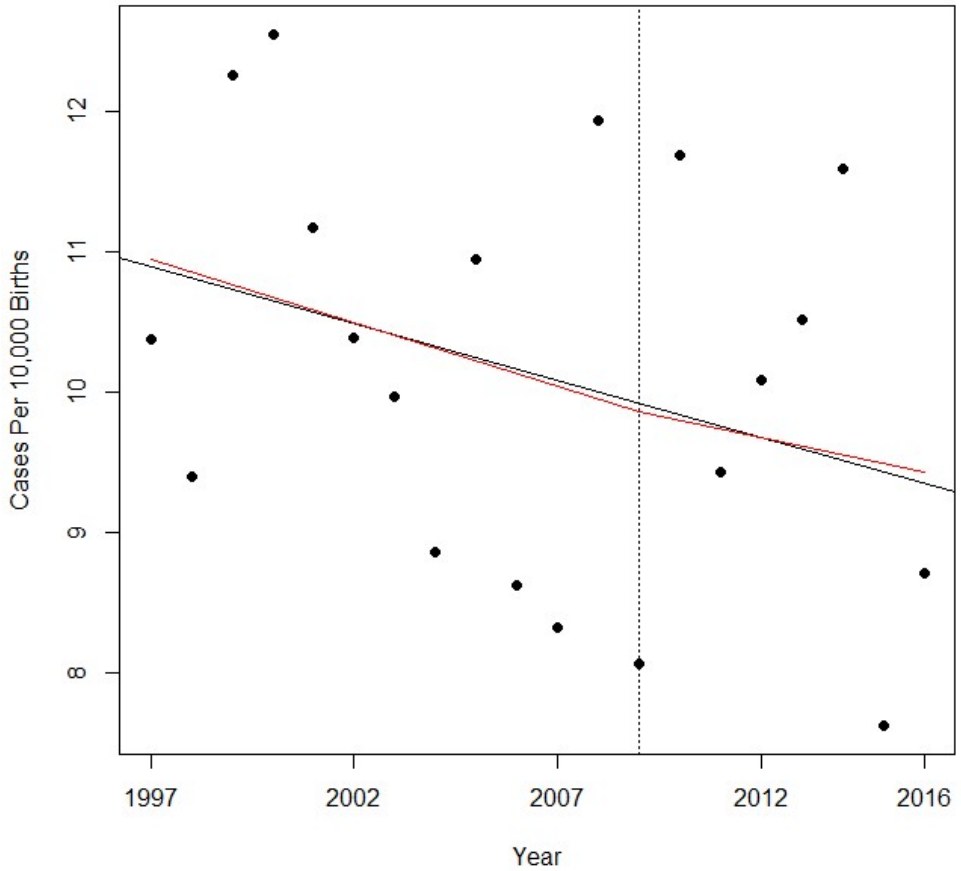
Piecewise Regression Model

- Simple linear regression: $\hat{Y} = \beta_0 + \beta_1 X$
- Piecewise regression model:

$$\hat{Y} = \begin{cases} \beta_0 + \beta_1 X & X \leq 2009 \\ (\beta_0 - 2009\beta_2) + (\beta_1 + \beta_2)X & X > 2009 \end{cases}$$

- Knot at $X = 2009$ selected by minimizing mean square error in the interval from 2005 to 2011

Model Comparison



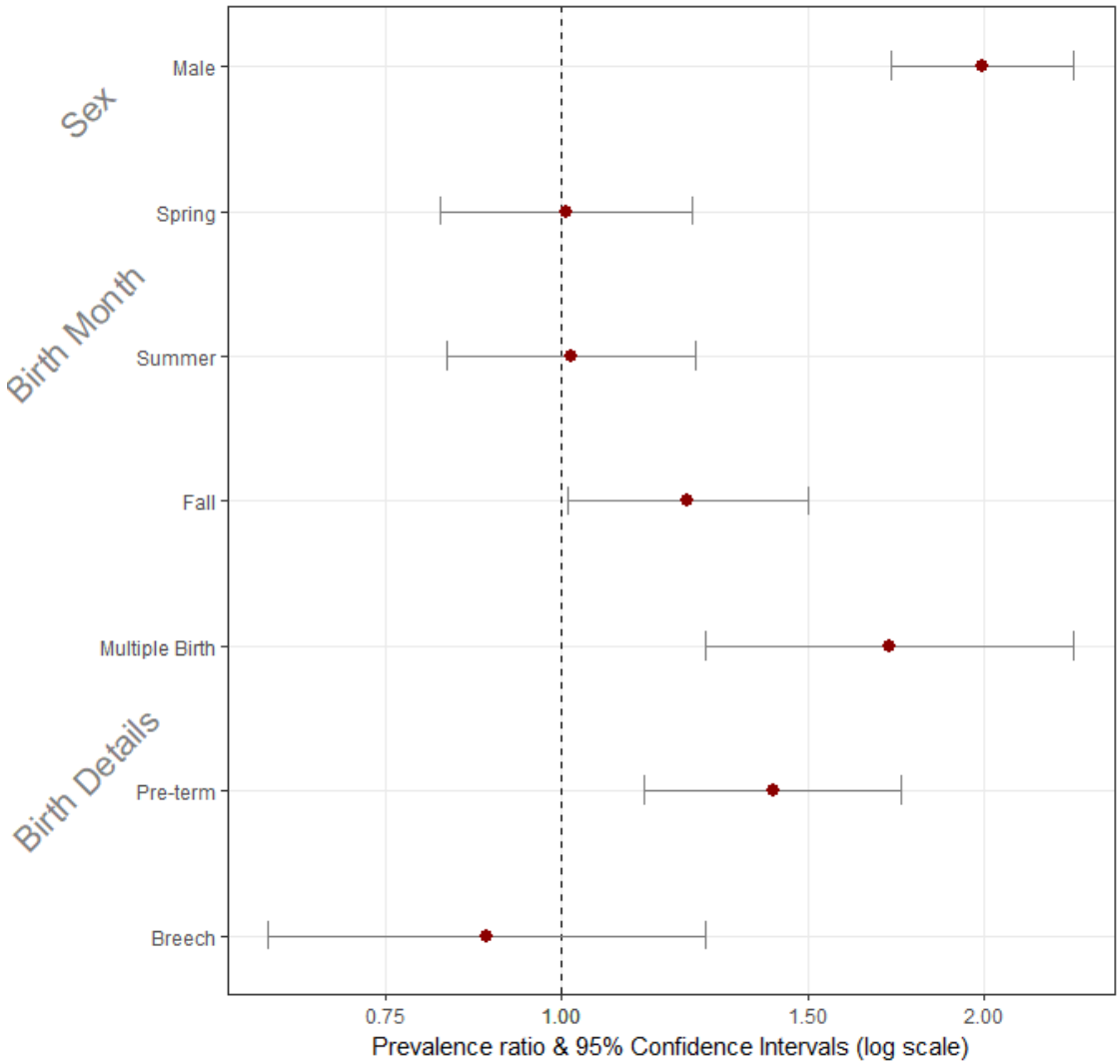
$$\beta_0 = 192.424, \beta_1 = -0.091, \beta_2 = 0.031$$

p-values: 0.337, 0.363, 0.904

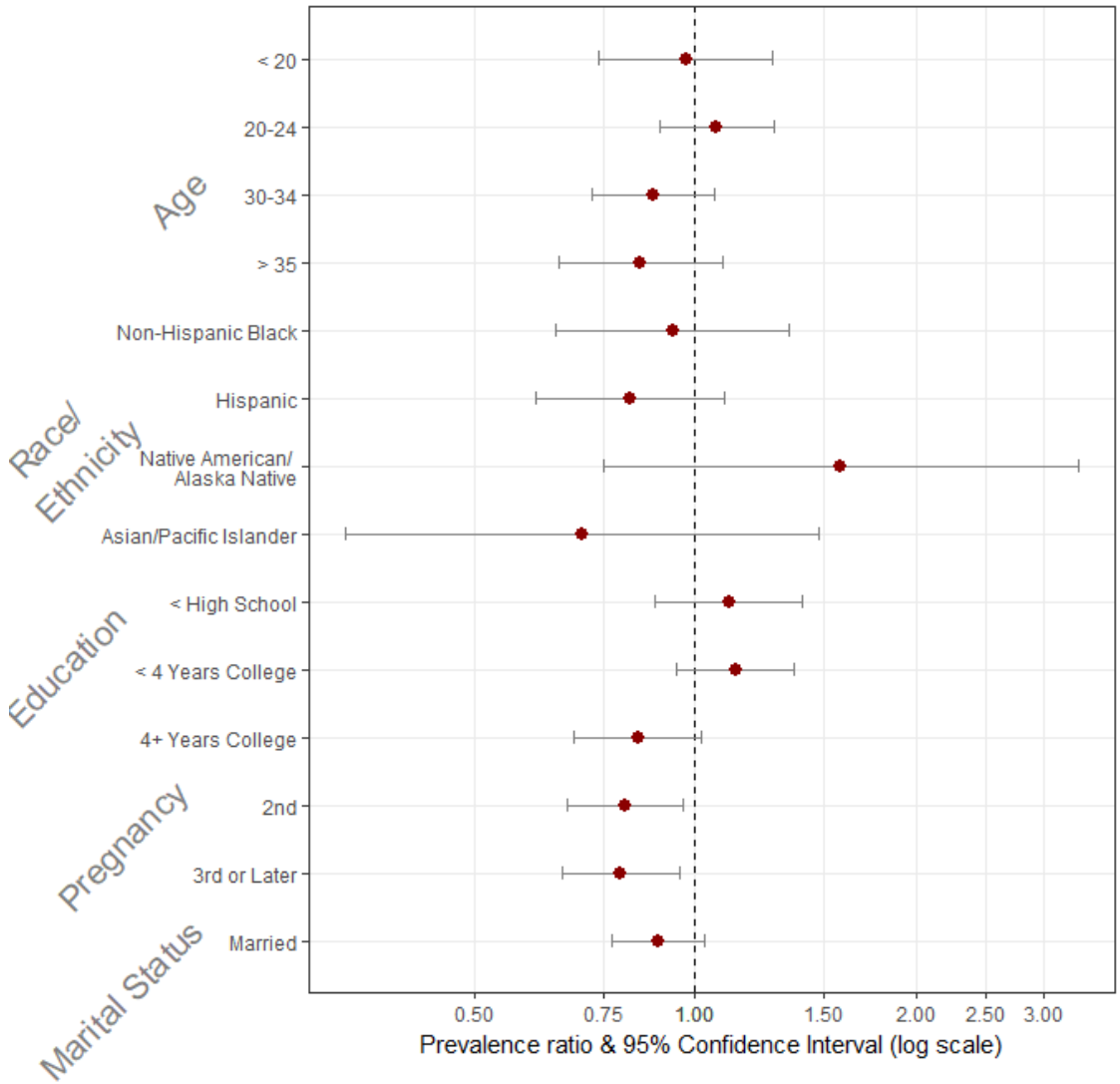
Identifying Risk Factors

- Selected characteristics available in IRCID and Iowa birth certificate data
- Logistic regression (appropriate due to binary outcome variable)
- Prevalence ratios & 95% confidence intervals calculated (using reference category/level for each factor)
- Cases to population comparison

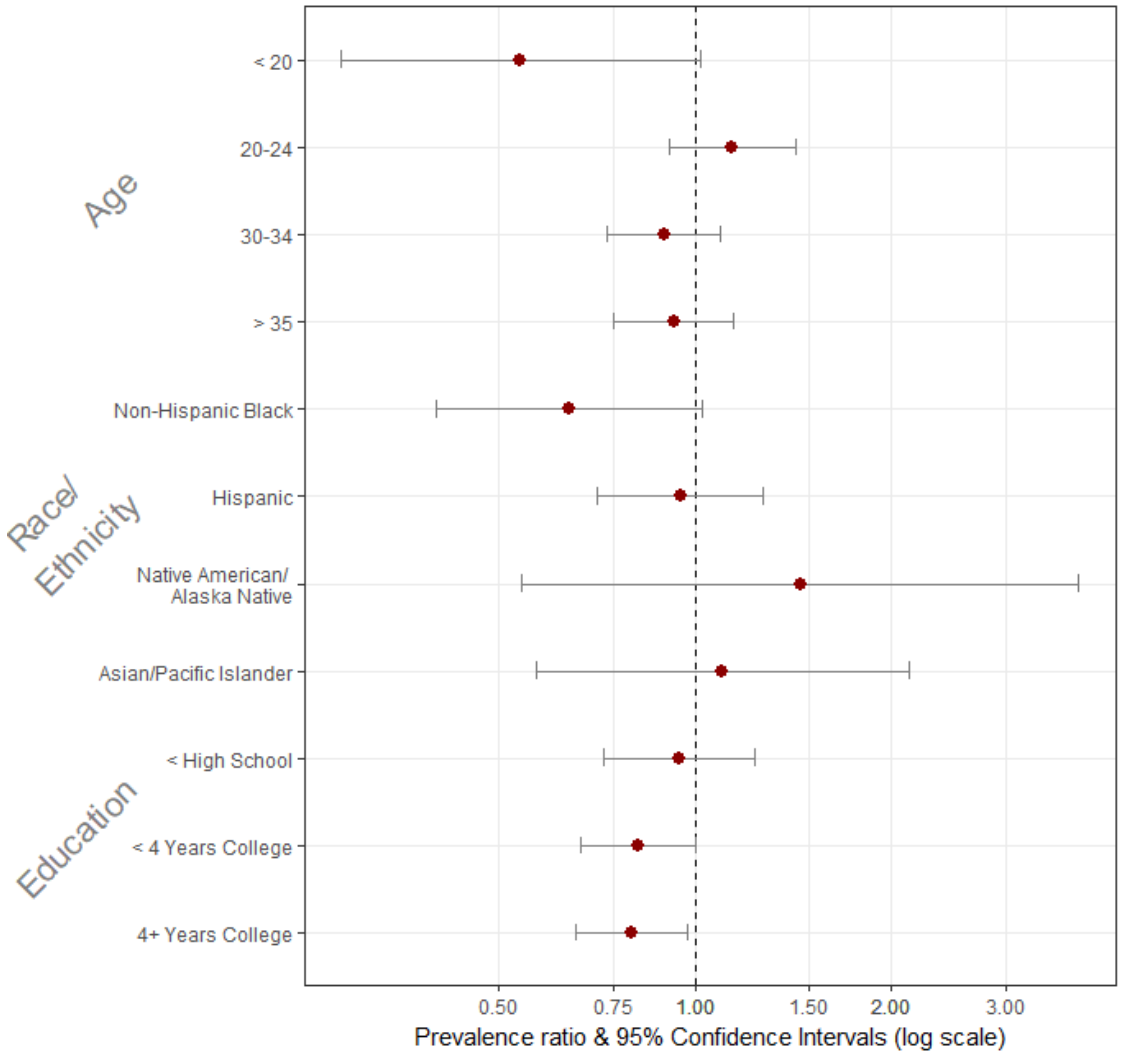
Child Characteristics



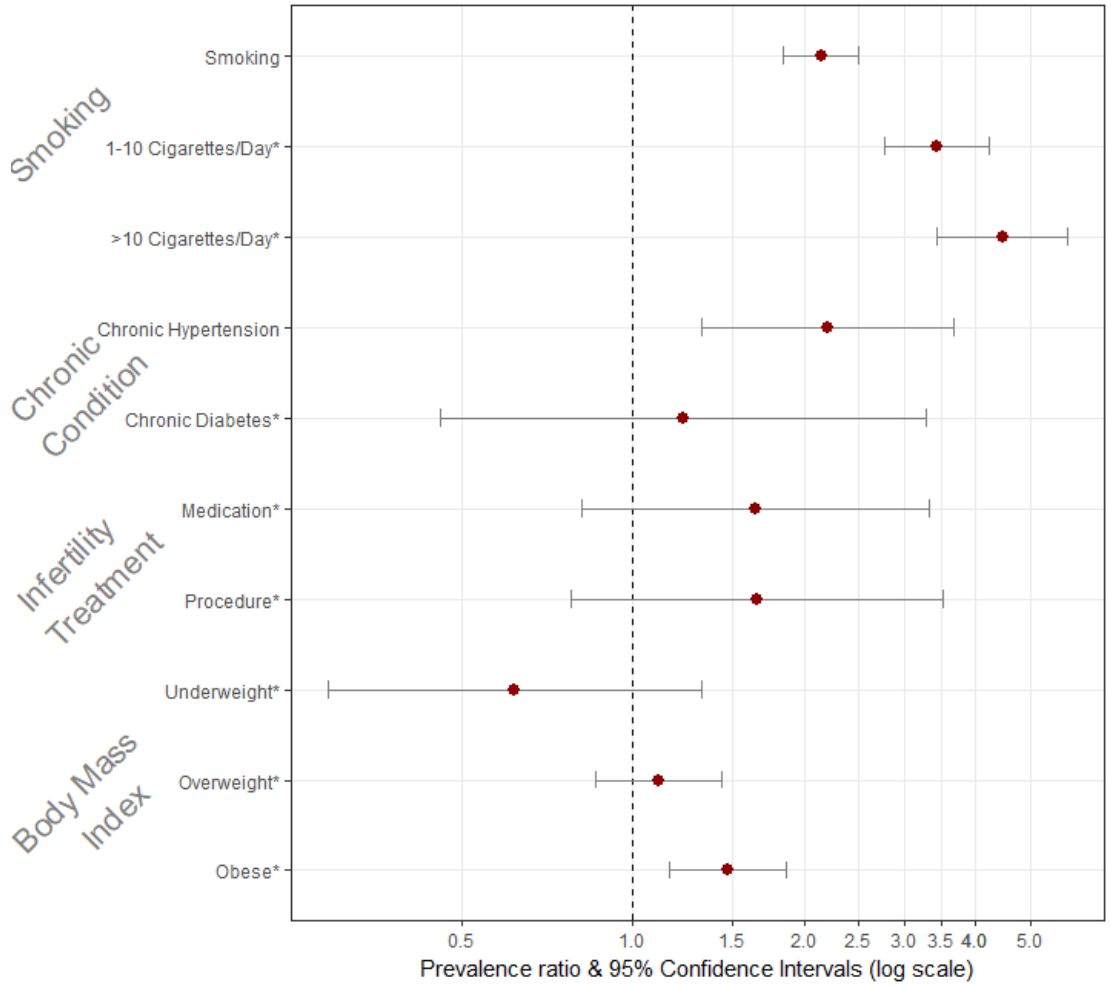
Maternal Characteristics



Paternal Characteristics



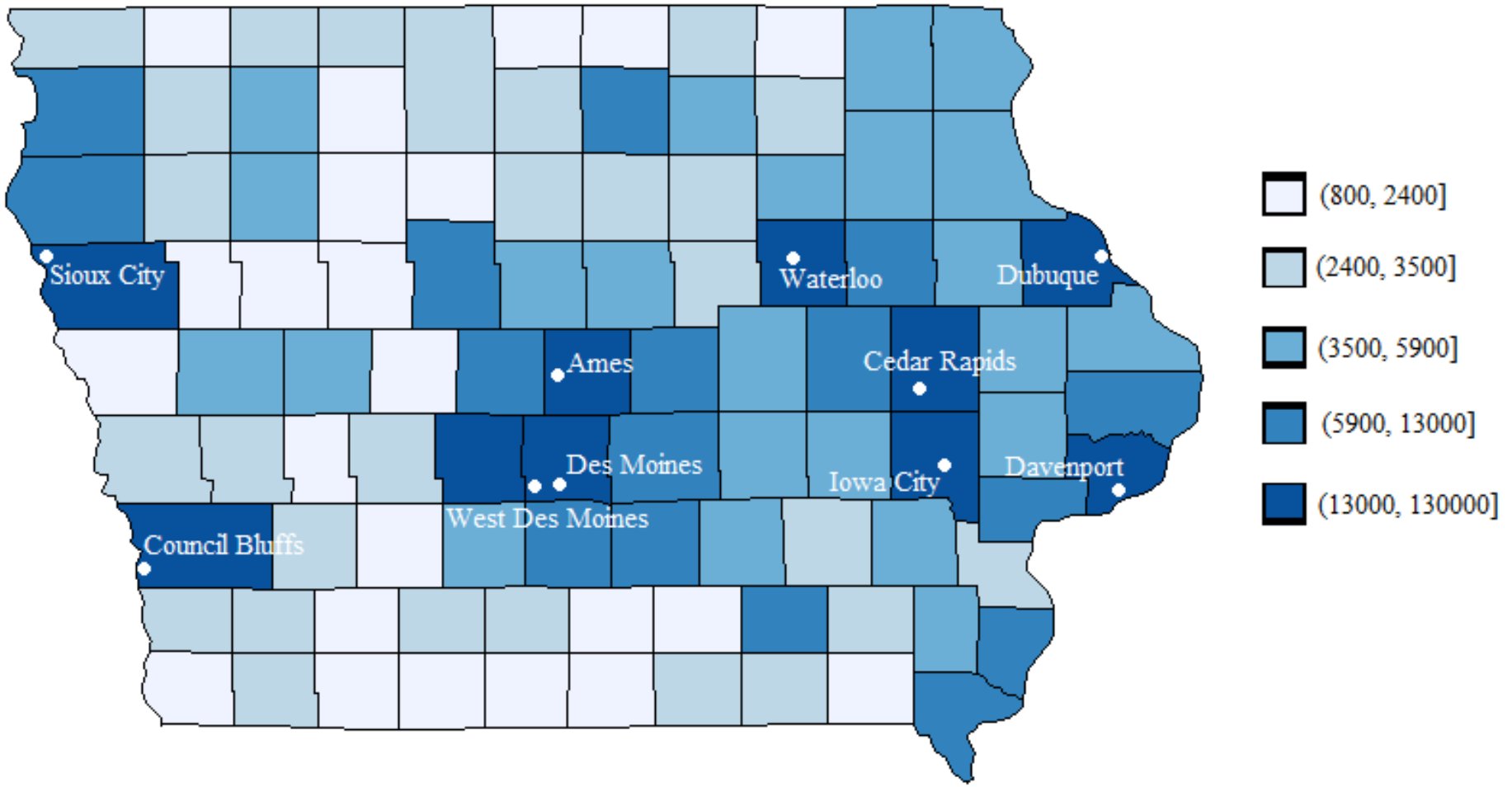
Maternal Exposures



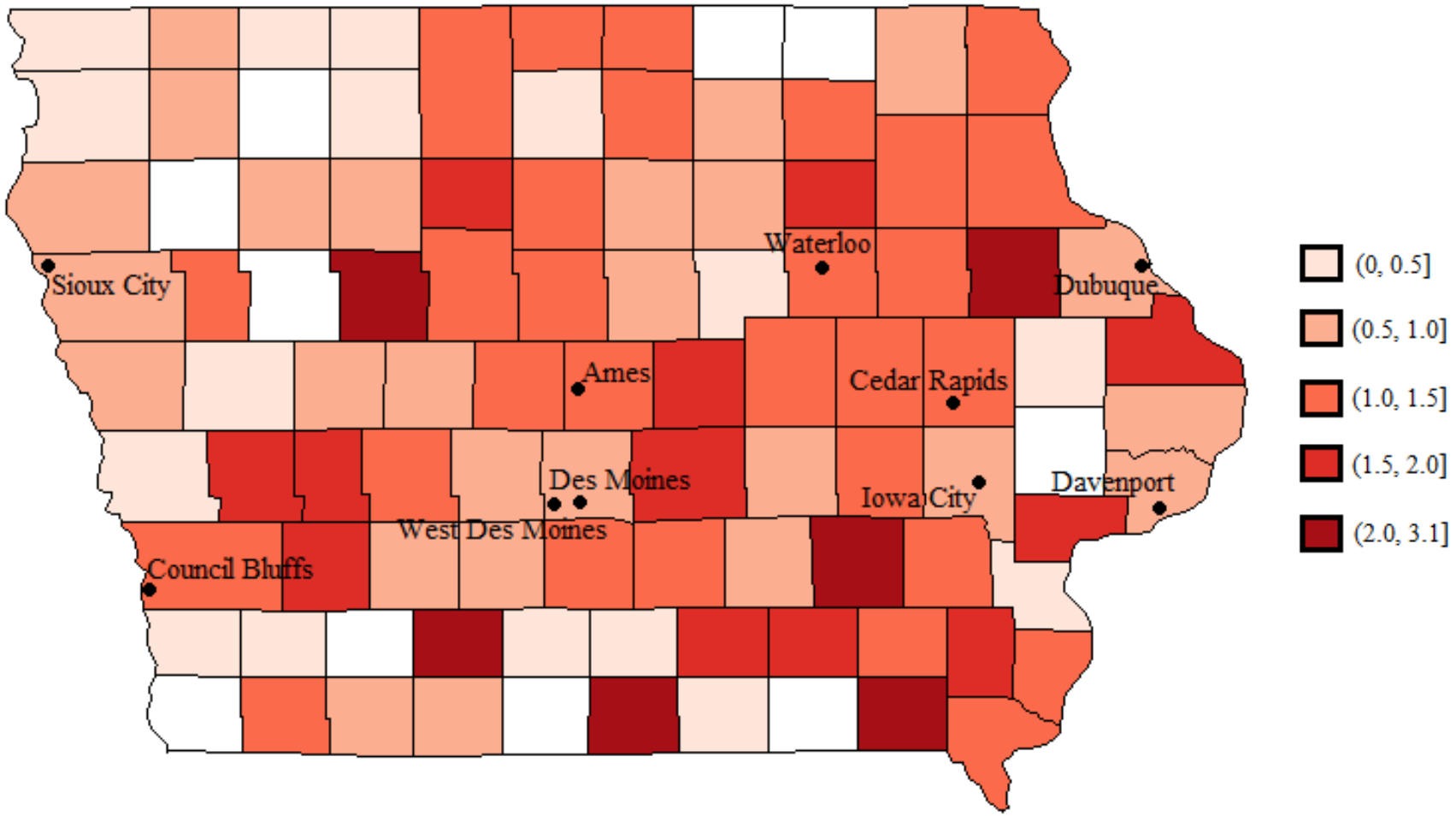
*Data for cigarettes/day, diabetes, infertility treatment & BMI only available for 2007-2016.

Spatial Analysis

Total Births by County



Raw Relative Risk by County



Poisson Regression Model

- Poisson GLM: $\log Y_{ij} = \beta_0 + \beta_1 X_j + \log E_{ij}$
- Y_{ij}: Observed count for county i in year j
- X_j: Year j
- E_{ij}: Expected count for county i in year j

Testing for Spatial Correlation

Moran's I Statistic

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- n: number of regions (99 counties)
- W: adjacency matrix ($W_{ij} = 1$ if county i and county j share a border; $W_{ii} = 0$)
- $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$
- x: outcome variable (observed clubfoot cases per county)

Testing for Spatial Correlation

Moran's I Test

- H_0: Observed counts are randomly distributed across counties; i.e. no spatial clustering of observed counts
- H_a: Observed counts are *not* randomly distributed

Testing for Spatial Correlation

Moran's I Test

- H_0: Observed counts are randomly distributed across counties; i.e. no spatial clustering of observed counts
- H_a: Observed counts are *not* randomly distributed
- Expected Moran's I under H_0: $E(I) = \frac{-1}{n-1} = -0.010$
- Calculated Moran's I: 0.091; p-value $= 0.013$

Testing for Spatial Correlation

Moran's I Test

- H_0: Observed counts are randomly distributed across counties; i.e. no spatial clustering of observed counts
- H_a: Observed counts are *not* randomly distributed
- Expected Moran's I under H_0: $E(I) = \dfrac{-1}{n-1} = -0.010$
- Calculated Moran's I: 0.091; p-value $= 0.013$
- *Conclusion*: There is spatial clustering at 5% significance level

CAR Prior for Spatial Random Effect

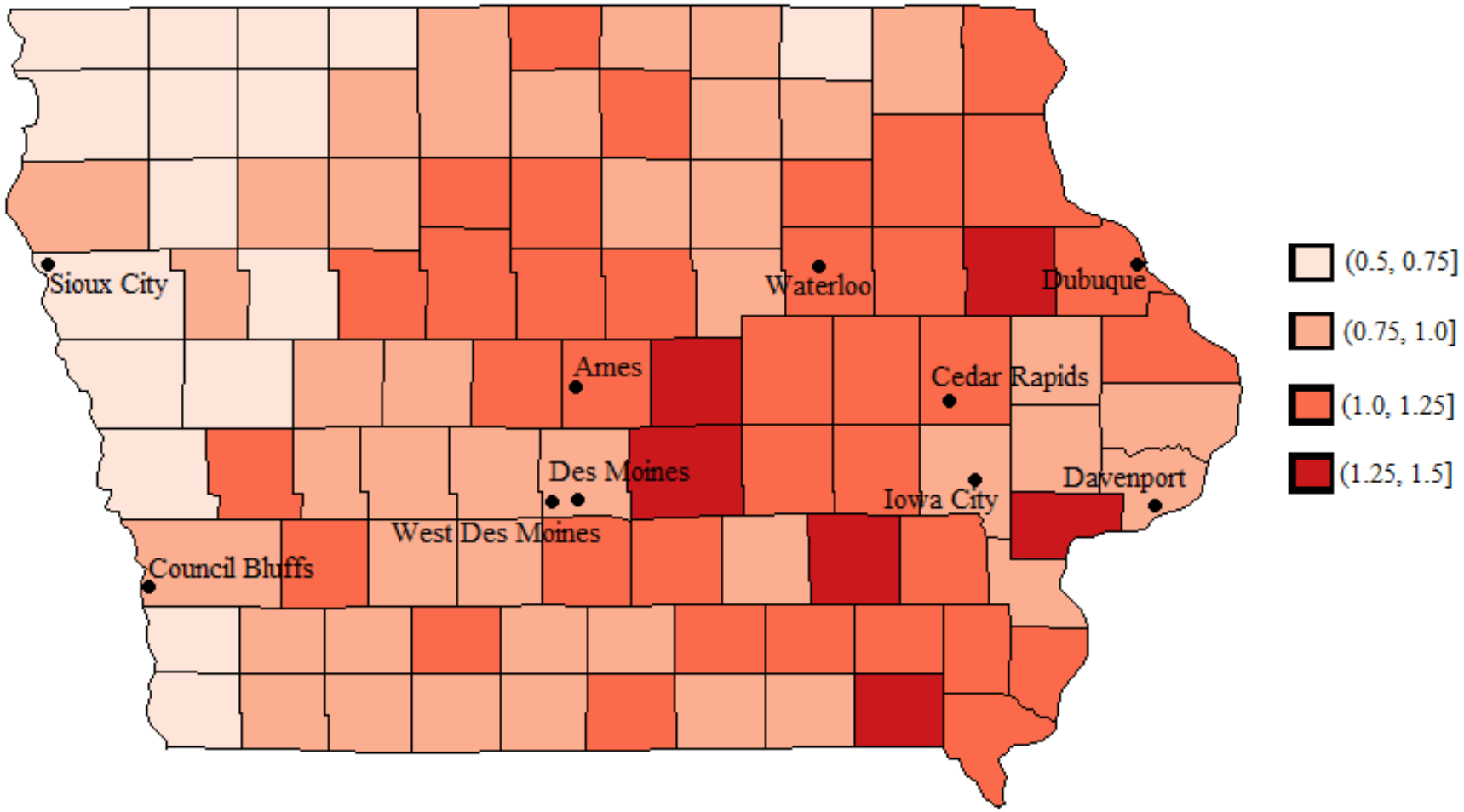
$$\phi \sim N(0, [\tau D - \alpha W]^{-1})$$

- $D = diag(m_i)$ (n x n diagonal matrix with m_i = number of neighbors of county i)
- W: adjacency matrix ($W_{ij} = 1$ if county i and county j share a border; $W_{ii} = 0$)
- α: binary spatial dependence parameter
- $\tau \sim$ Gamma(2,2)
- Bayesian approach implemented using **rstan** package in R

Poisson Model with Spatial Error Factor

- $\log\ Y_i = \beta X_i + \phi_i + \log\ E_i$
- $X_i = 1\ \forall\ i$ (no other covariates included)

Relative Risk by County (Spatially Smoothed)



Conclusions

- Prevalence relatively constant over 20-year time period
- Potential risk factors:
  - Maternal smoking, increased BMI, hypertension
  - Multiple birth, preterm birth
  - Lower socioeconomic status
- Possible hotspots:
  - Keokuk County
  - Delaware County
  - Van Buren County

Acknowledgements

- Jake Oleson
- Paul Romitti
- Anthony Rhoads & Florence Foo
- Gideon Zamba
- Grant Brown
- Monica Ahrens & Javi Flores (and many other graduate students)
- Department of Biostatistics
- ISIB 2018 Cohort (special thanks to Isaac Slagel)
- NHLBI (Grant #HL131467)

References

Joseph, Max. "Exact Sparse CAR Models in Stan." August 20, 2016. http://mc-stan.org/users/documentation/case-studies/mbjoseph-CARStan.html.

Kancherla V, Romitti PA, Caspers KM, Puzhankara S, Morcuende JA. 2010. Epidemiology of congenital idiopathic talipes equinovarus in Iowa, 1997–2005. Am J Med Genet Part A 152A:1695–1700.

Werler MM, Yazdy MM, Mitchell AA, et al. Descriptive Epidemiology of Idiopathic Clubfoot. American journal of medical genetics Part A. 2013;161(7):1569-1578.