An Evaluation of Elevated Blood Lead Levels in Iowa's Newborns

By: Helin Hernandez and Charimhel Rarallo
Faculty Mentors: Jacob Oleson and David Zahrieh
Iowa Summer Institute in Biostatistics 2016

What is Lead?

- A natural metal found deep in the ground and in small quantities all throughout our planet
- It can also be found in other elements
- Dangerous because it has no small and it's hard to detect
- Was used in the manufacturing of many products






History and Today

- 1922: Lead usage peaked
    - Washable and durable
    - Scientific studies were conducted to check safety of lead
        - Raised concerns
- 1940s and 1950s: Use of lead decreased
- 1971: Lead Poisoning Act was passed
    - CDC established Blood Lead Level (BLL) safety threshold of 60 micrograms/deciliter
- 1978: Lead products completely banned
    - Lead lingered
- 1991: CDC lowered BLL to 10 mg/dL
- 2012 : New BLL threshold of 5 mg/dL
    - No known "SAFE" BLL

Consequences of Lead

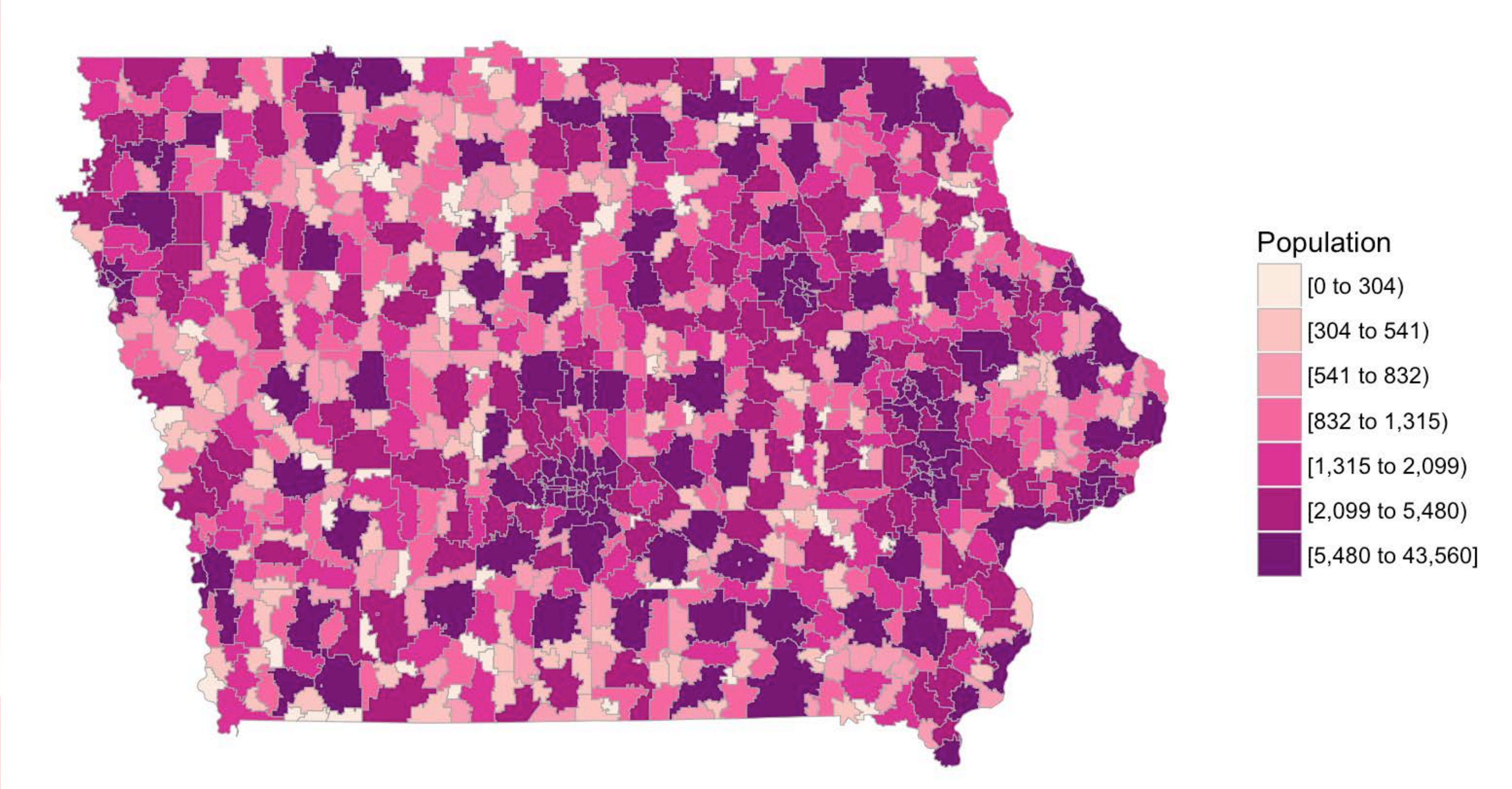
- How does lead get into our bodies?
  - Respiratory and Digestive systems
- Consequences of lead exposure
  - No unique symptoms
  - Health risks: headaches, stomach pain, body development issues, miscarriages and premature birth and more
- Flint, Michigan example
  - About half Flint's water source pipings are made of lead
  - Residents suffered from health problems named above

Research Interest

Locate areas within Iowa with greater risk of exposure to high lead levels

- U.S. Census Bureau Data
  - 5 covariates of interest
  - ZCTA (ZIP Code Tabulated Areas)
- Region where the newborn's mother lived prior to giving birth
  - ZIP Code
- Response variable (Tally) measures the number of newborns with elevated BLL in each ZCTA
- Offset variable is the number of women of children bearing age per ZCTA

2012 ZIP code Population Estimates



U.S. Census Bureau Data

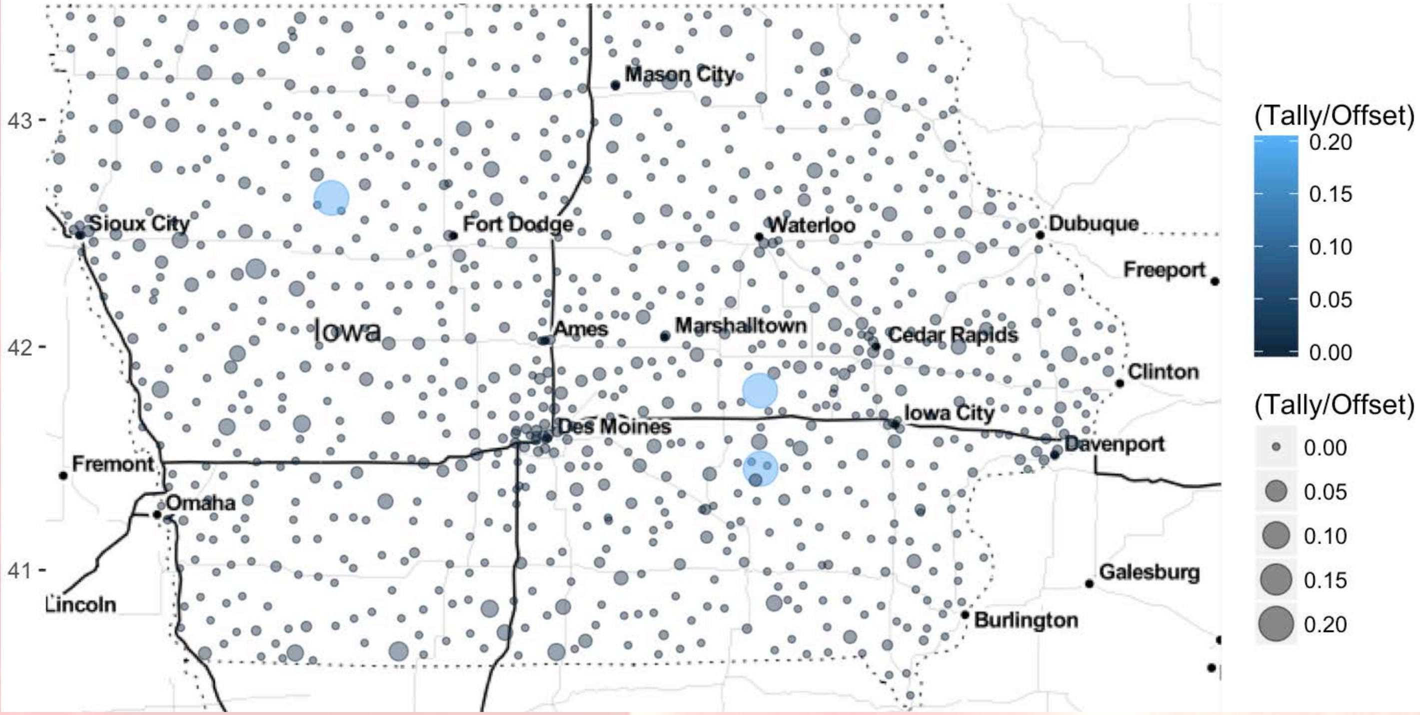
- Annual American Community Survey (ACS)
- 2007- 2011 variable means
- Variables used (by ZCTA):
  - **pre40:** Number of homes built before 1940
    - Lead used in household paint
  - **med_inc:** Median income in household
  - **edu_ltco:** Number of females of child bearing age with less than a college education
  - **fem_for:** number of female foreigners
  - **fem_pov:** number of females of child bearing age living in poverty

Collected Data

- Evaluated 2741 Iowan newborn dried blood spots
- Samples identified by mother’s ZIP code of residence
- 2006, 5 months, Iowa

- Tally: The number of newborns with elevated BLL per ZCTA
- Offset: The number of women of child bearing age per ZCTA

Ratio of Population with Elevated BLL



# of Newborns with High BLL	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15
Frequency	238	132	53	21	14	7	8	2	1	4	3	4	2	2	1

Data Analysis: Missing Data

- Some of the Covariates had missing data in certain ZCTAs.

	pre40	med_inc	edu_ltco	fem_for	fem_pov
# of NAs	5.00	7.00	1.00	1.00	5.00
% of NAs	0.01	0.01	0.00	0.00	0.01

- Imputation of NAs in the five covariates of interest
- Multivariate Imputation by Chained Equation (MICE) method

Data Analysis: Spatial Correlation

- Spatial relationship between high lead levels and region of birth
  - Moran's Test for spatial autocorrelation using a spatial weights matrix in a list structure
    - H_0 : No spatial correlation between ZCTA
    - H_a : Spatial correlation between ZCTA
      - P-value: 2.2e-16
      - Moran's Statistic: 0.2242

Poisson Regression

Goal: to evaluate the relationship between the number of babies with elevated blood lead levels to the five ZIP code level covariates of interest

- Poisson Regression approach, without spatial component, was our first attempt because of limited time
- Poisson Regression deals with discrete outcome variables that occur within a specific time and place.
  - $Y_i \sim Poisson(\theta_i)$
- Akaike Information Criterion (AIC) was used to evaluate the accuracy of each model
- Generalized linear model (glm) command was used specifying family = "Poisson"
- The model below was the model with the lowest AIC

$$\log(\theta_i) = \beta_0 + \beta_1(pre40) + \beta_3(edultco) + \beta_5(fempov) + \beta_6(edultco{:}\,fempov) + \log(\,Offset)$$

- 2 covariates and 1 interaction as significant

Bayesian Poisson Regression

$$\log(\theta_i) = \beta_0 + \beta_1(pre40) + \beta_3(edultco) + \beta_5(fempov) + \beta_6(edultco\!:fempov) + log(\,Offset) + Z_i$$

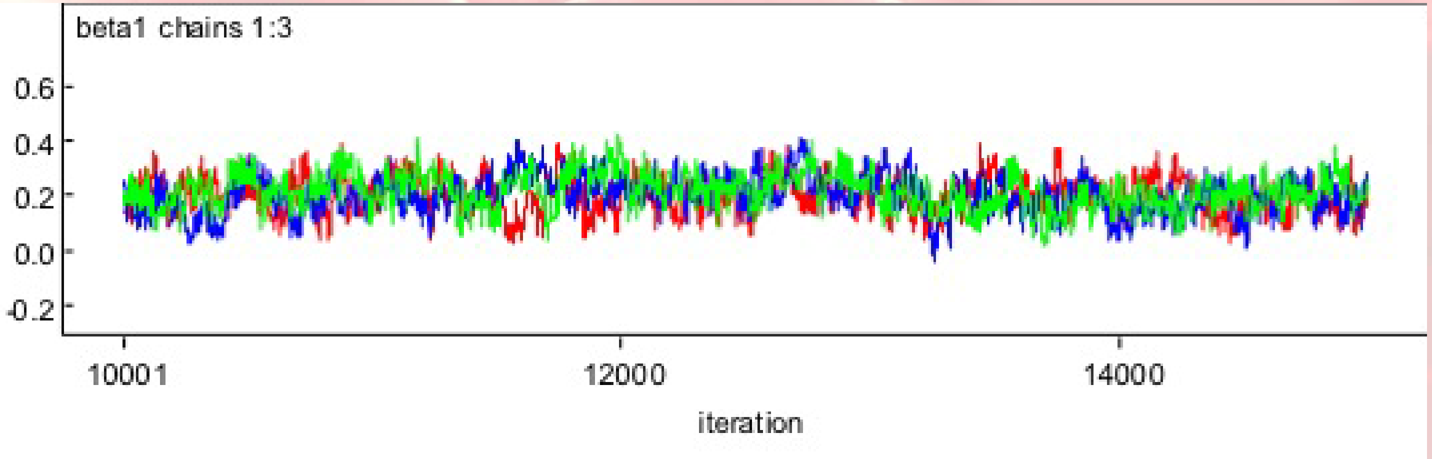
- $\beta_0 \sim N(0, 1000)$
- $\beta_1 \sim N(0, 1000)$
- $\beta_3 \sim N(0, 1000)$
- $\beta_5 \sim N(0, 1000)$
- $\beta_6 \sim N(0, 1000)$

- $Z_i \sim Conditional\ Auto\ Regression\ (CAR)$
  - $\sigma^2 \sim InverseGamma(0.01, 0.01)$

Bayesian analysis essentially returns a weighted average based on each ZCTA's mean lead level value and the mean lead level values of it's neighboring ZCTA.

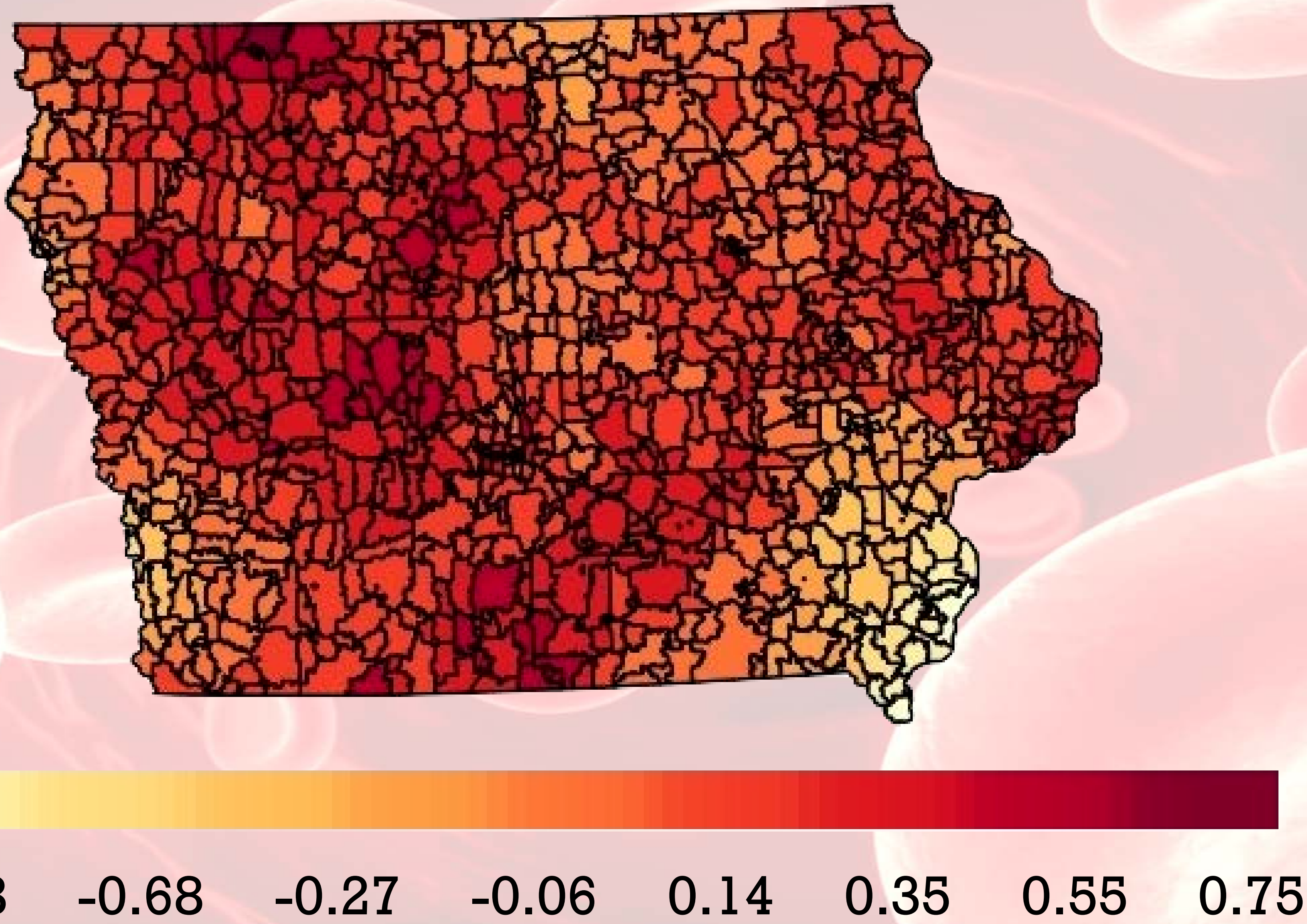
Bayesian Generalized Linear Mixed Model

- Utilized R2OpenBUGS to run the Bayesian Poisson model through R
- 3 different start points
- Burn-in 10,000 and sample the next 5,000
- Only Pre40 covariate (β_1) was significant
- Education variable (β_3) and the intersection between education and poverty variables (β_6) just barely encompassed zero in the 95% credible interval



	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-6.66	0.18	-37.34	0.00
pre40	0.01	0.00	3.57	0.00
edu_ltco	-0.00	0.01	-0.35	0.73
fem_pov	-0.03	0.01	-3.60	0.00
edu_ltco:fem_pov	0.00	0.00	2.38	0.02

$$\log(\theta_i) = \beta_0 + \beta_1(pre40) + \beta_3(edultco) + \beta_5(fempov) + \beta_6(edultco{:}\,fempov) + log(\,Offset) + \boldsymbol{Z_i}$$



Conclusion

- Bayesian Poisson regression showed different results than our traditional Poisson regression
- Bayesian models better accept missing values in response variables
  - Tally was 47% missing
  - Considered spatial correlation as opposed to the traditional Poisson regression
  - Education covariate and interaction term between education and poverty variables were almost significant

Concerns

- 2 months vs. 5 months
- Analyzed sample dates: some months had no results
- Only able to consider 2-factor interactions due to time limit

Limitations

- Missing data
- How long the analysis took to run
    - 2134 sec ≈ 36 mins
    - 15,000 iterations: 5,000 each
- Software problems: Do not be a MAC user ☹

Bibliography

- Childhood Lead Poisoning Data, Statistics, and Surveillance. (2016, March 25). Retrieved July 16, 2016, from http://www.cdc.gov/nceh/lead/data/index.htm
- Gains, S., & Tran, L. (2016, January 13). How Flint, Michigan's tap water became toxic. Retrieved July 18, 2016, from http://www.cnn.com/2016/01/11/health/toxic-tap-water-flint-michigan/
- Lead. (n.d.). Retrieved July 13, 2016, from http://www.niehs.nih.gov/health/topics/agents/lead/
- Lead Poisoning. (n.d.). Retrieved July 13, 2016, from http://kidshealth.org/en/parents/lead-poisoning.html
- Lead Poisoning. (2016). Retrieved July 16, 2016, from http://www.cdc.gov/nceh/lead/
- Prevention Tips. (2014). Retrieved July 13, 2016, from https://www.cdc.gov/nceh/lead/tips.htm
- The History of the Use of Lead-Based Paint - Understanding Lead Pigment Litigation. (n.d.). Retrieved July 13, 2016, from http://www.leadlawsuits.com/history/history-of-the-use-of-lead-paint/

Thank You

National Institute of Health

University of Iowa, Biostatistics Department

Faculty Mentor Jacob Oleson

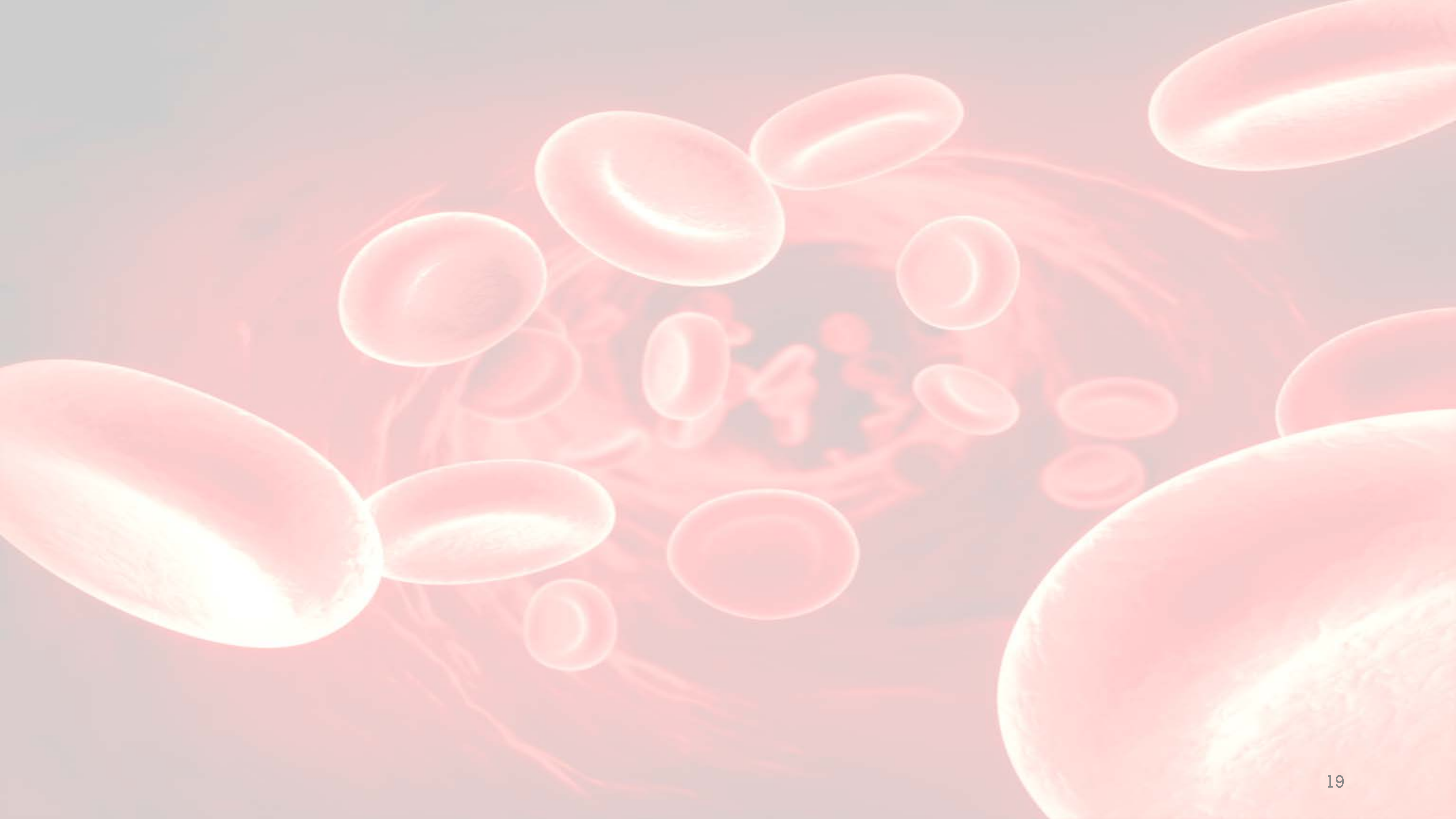
Graduate Student Mentor David Zahrieh

Special Thanks to:

Lauren Sagar and Grant Brown for Technical support

ISIB Program sponsored by the National Heart Lung and Blood Institute (NHLBI) HL131467

Questions? ☺



Multivariate Imputation by Chained Equation (MICE) method

1. $x_1 \sim x_2 + x_3 + x_4$

Or

1. $x_1^* \sim E(x_2) + x_3 + x_4$
2. $x_2^* \sim x_1^* + x_3 + x_4$
3. $x_1^{**} \sim x_2^* + x_3 + x_4$

.

.

.

(repeat 10 times and use the 10^{th} prediction for each missing value)

Spatial Variable in Bayesian Poisson Model

$$Z_i \sim N(0\,,\delta(I\ - C)^{-1})$$

$$\delta \sim InverseGamma\,(0.01, 0.01)$$
$$C = ZCTA\,Neighborhood\,Matrix$$

Coefficient Estimates for both Regression Analysis

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-6.66	0.18	-37.34	0.00
pre40	0.01	0.00	3.57	0.00
edu_ltco	-0.00	0.01	-0.35	0.73
fem_pov	-0.03	0.01	-3.60	0.00
edu_ltco:fem_pov	0.00	0.00	2.38	0.02

Traditional Poisson Regression

Covariate	Mean	SD	P_2.5	P_97.5
Intercept	-6.43	0.07	-6.56	-6.30
pre40	0.20	0.07	0.08	0.33
edu_ltco	0.14	0.09	-0.02	0.32
fem_pov	-0.06	0.07	-0.20	0.07
edu_lcto:fem_pov	0.11	0.07	-0.02	0.23

Bayesian Poisson Regression