Spatial analyses of county-level birth and death data from the National Vital Statistics System

Lauren M. Rossen, PhD, MS

Office of Analysis and Epidemiology
National Center for Health Statistics

Geospatial Statistics, Tools, Data, Practices, Opportunities and Challenges in the Federal Agencies

October 16, 2015

Acknowledgements

Co-authors:

Diba Khan
Brady Hamilton
Margy Warner

Colleagues in the Division of Vital Statistics, Office of Research Methodology, and the Research Data Center

Spatial Analyses of Birth and Death Data

Examples:

1. Drug Poisoning Death Rates in the U.S., 2002-2013
   - Two-stage hierarchical generalized linear models

2. Teen Birth Rates in the U.S., 2003-2012
   - Hierarchical Bayesian space-time interaction models

First Example

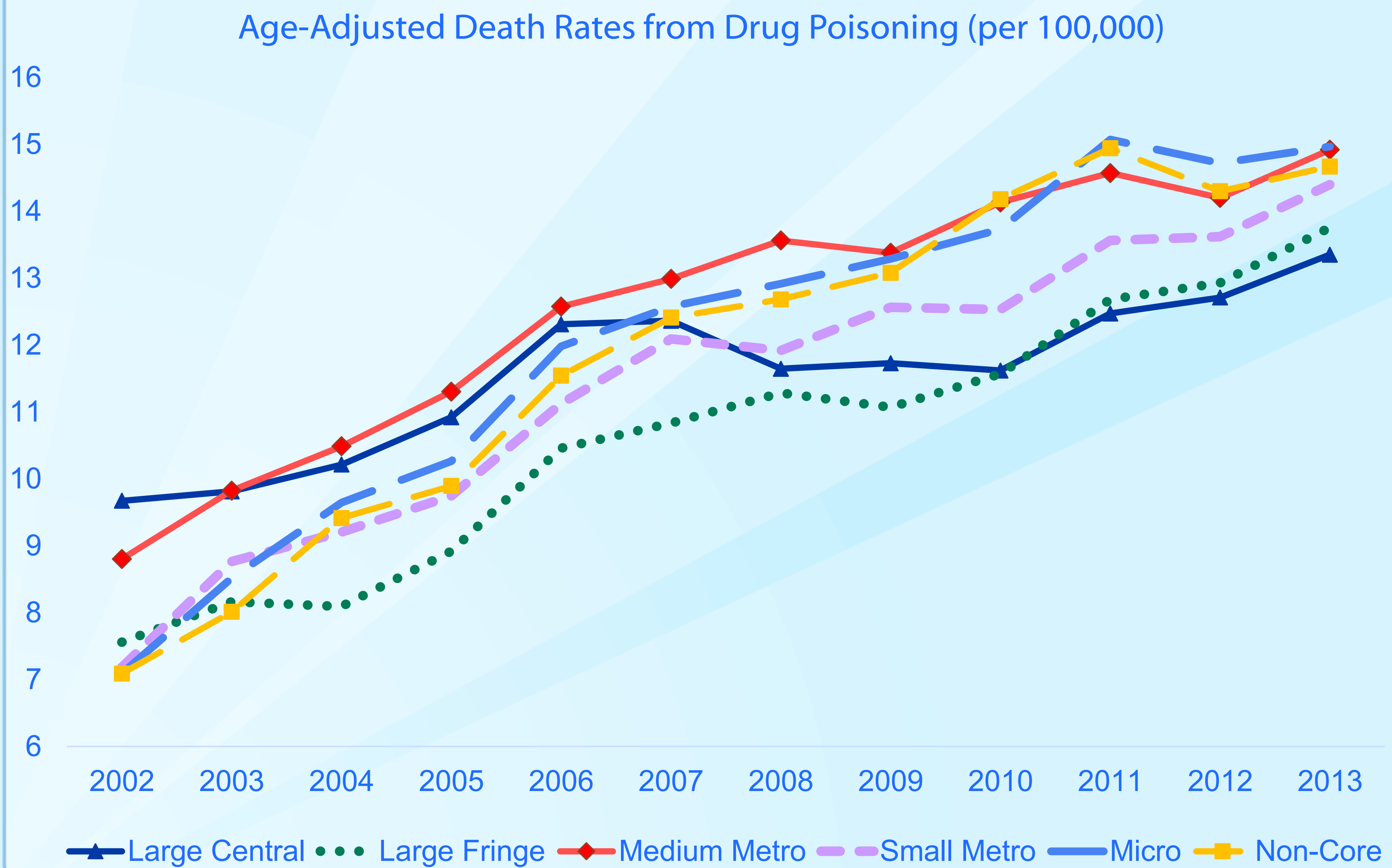
Drug Poisoning Mortality, 2002-2013

Drug Poisoning Mortality, 2002-2013

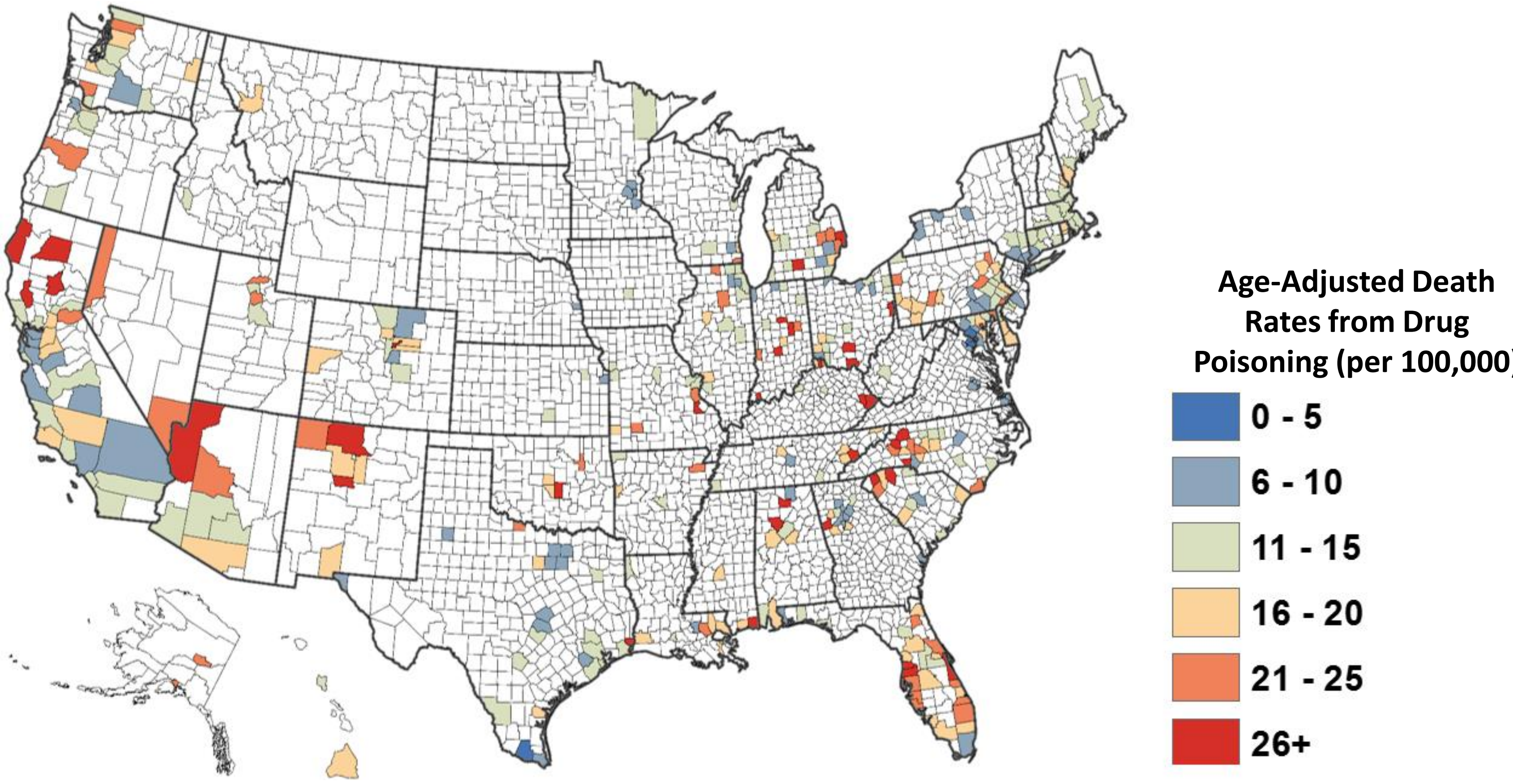
BACKGROUND

- Death rates associated with drug poisoning have doubled since 2000, to ~ 14 per 100,000 in 2013
  - More deaths due to drug poisoning than motor vehicle crashes
  - Drug overdoses are a major public health concern
- Death rates highest in West Virginia (32), Kentucky (24), New Mexico (23), Rhode Island (22) and Utah (22)
- Interest in county-level variation:
  - Where are death rates due to drug poisoning highest or lowest?
  - Where have we seen larger or smaller increases over time?

Trends by Urban-Rural Designation

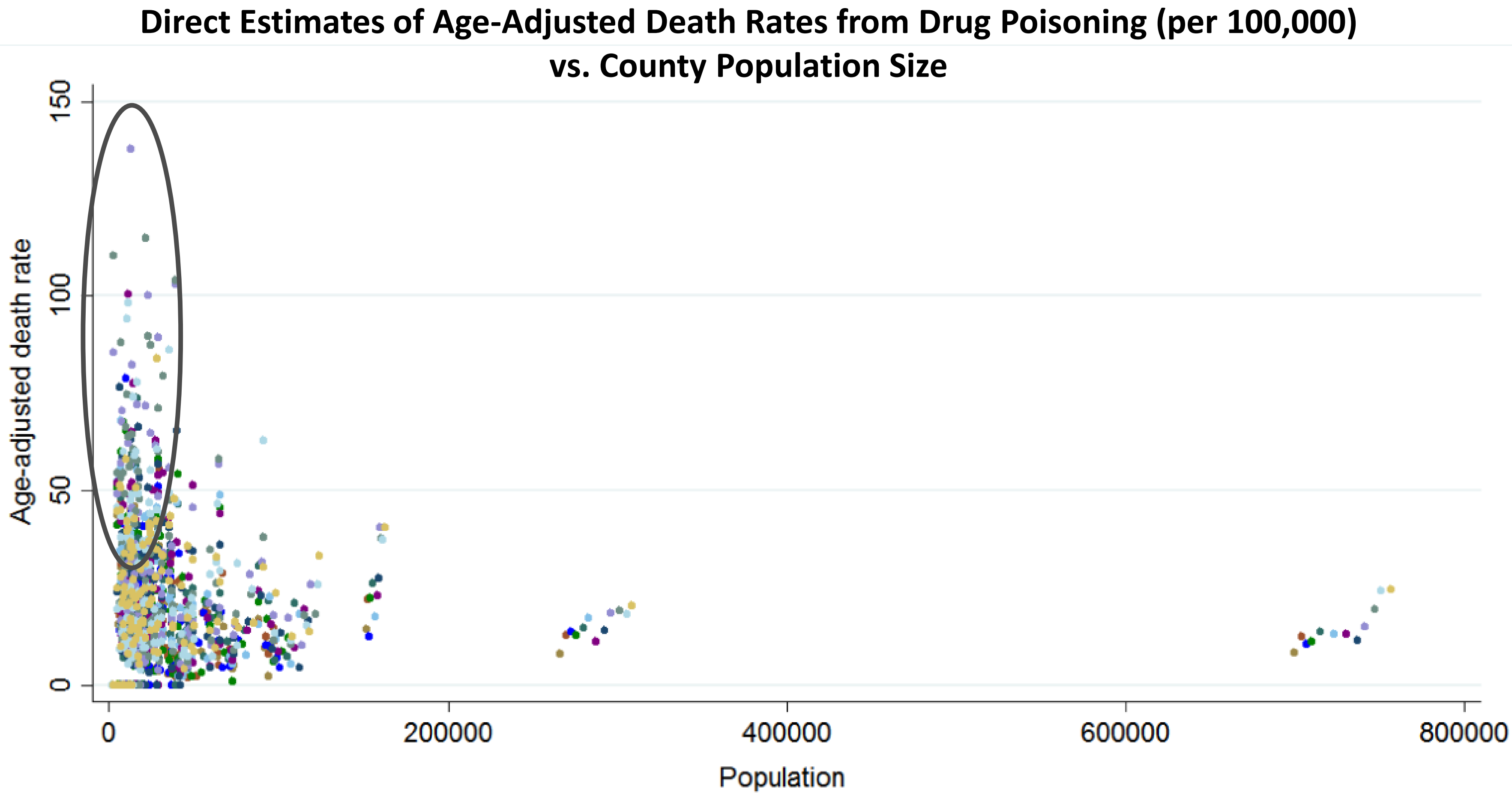


RATIONALE FOR SMOOTHING



- Death rates with data suppressed for counties with < 20 deaths in 2009
  - ~ 87% of counties suppressed!
  - Rare outcomes ➔ cannot look at sub-state variation using direct estimates

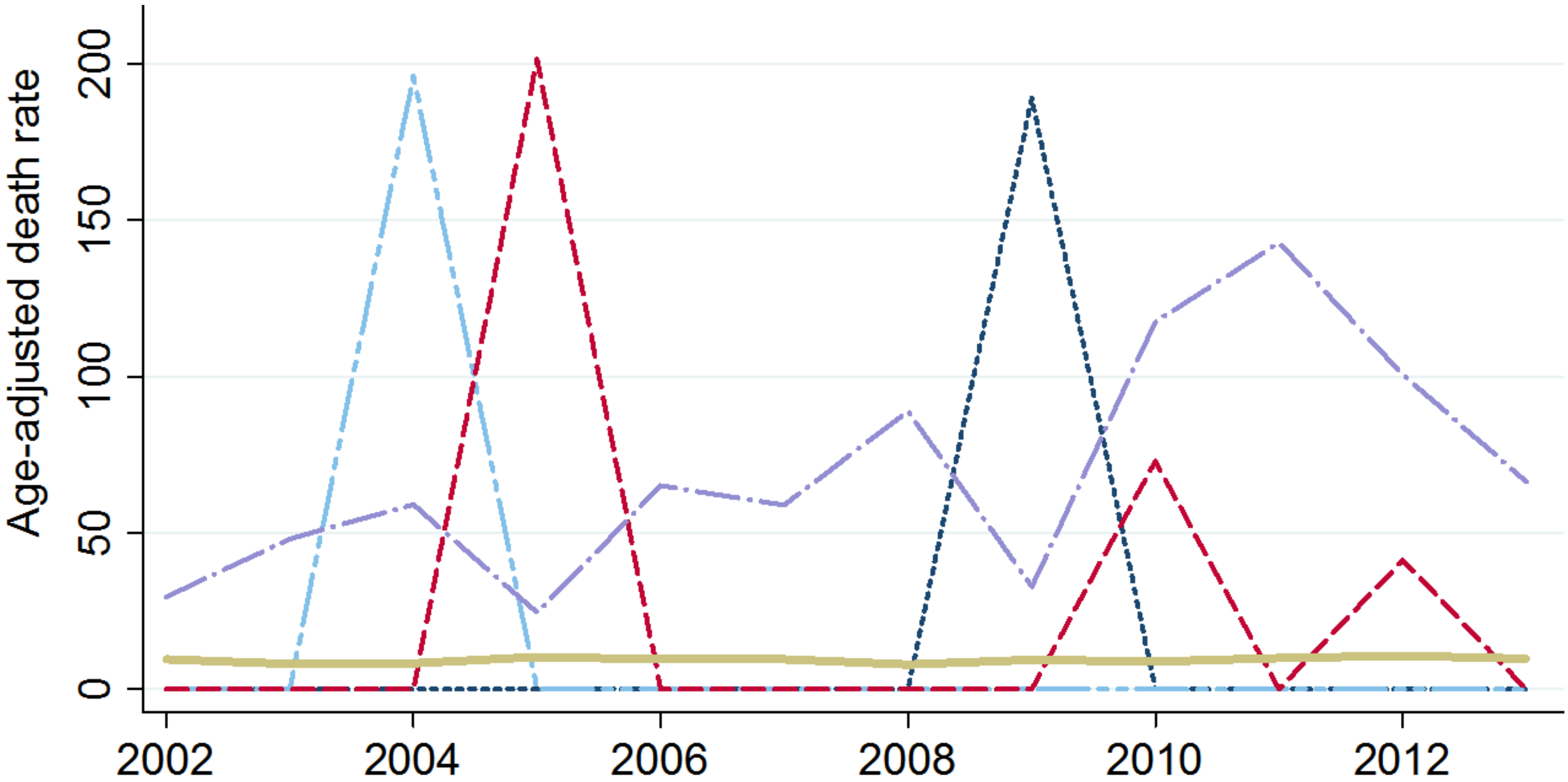
RATIONALE FOR SMOOTHING (continued)



- Rates are unstable for counties with small populations
    - Could combine years, but may mask temporal trends

AN EXAMPLE OF UNSTABLE RATES…

Direct Estimates of Age-Adjusted Death Rates from Drug Poisoning (per 100,000)



- Solid sand-colored line is a large city, other 4 counties are small
  - Death rates fluctuate from 0 to 200 per 100,000 year-to-year

DATA AND ANALYSES

- y_{it} = Age-adjusted death rate (AADR) from drug poisoning for county *i* at time *t*
  - from National Vital Statistics Multiple Cause of Death Files, 2002-2013
- y_{it} ~ highly zero-inflated, right-skewed distribution
  - Use two-stage models
    - » Stage 1: model probability of observing a death
    - » Stage 2: model death rate, given death was recorded

TWO STAGE MODELS

Stage 1: $\text{logit}(y_{it}=0) = \alpha^{(1)} + A_i^{(1)} + B_t^{(1)} + X_i'\gamma^{(1)}$

Stage 2: $\log(y_{it}|y_{it}>0) = \alpha^{(2)} + A_i^{(2)} + B_t^{(2)} + X_i'\gamma^{(2)}$

α = intercept

A_i = county-level random effect

B_t = fixed effects for year

$X_i'\gamma$ = vector of covariates and corresponding parameters, γ

- urban/rural classification
- socio-demographic characteristics at the county-level
- economic characteristics at the county-level

SMOOTHED COUNTY-LEVEL ESTIMATES

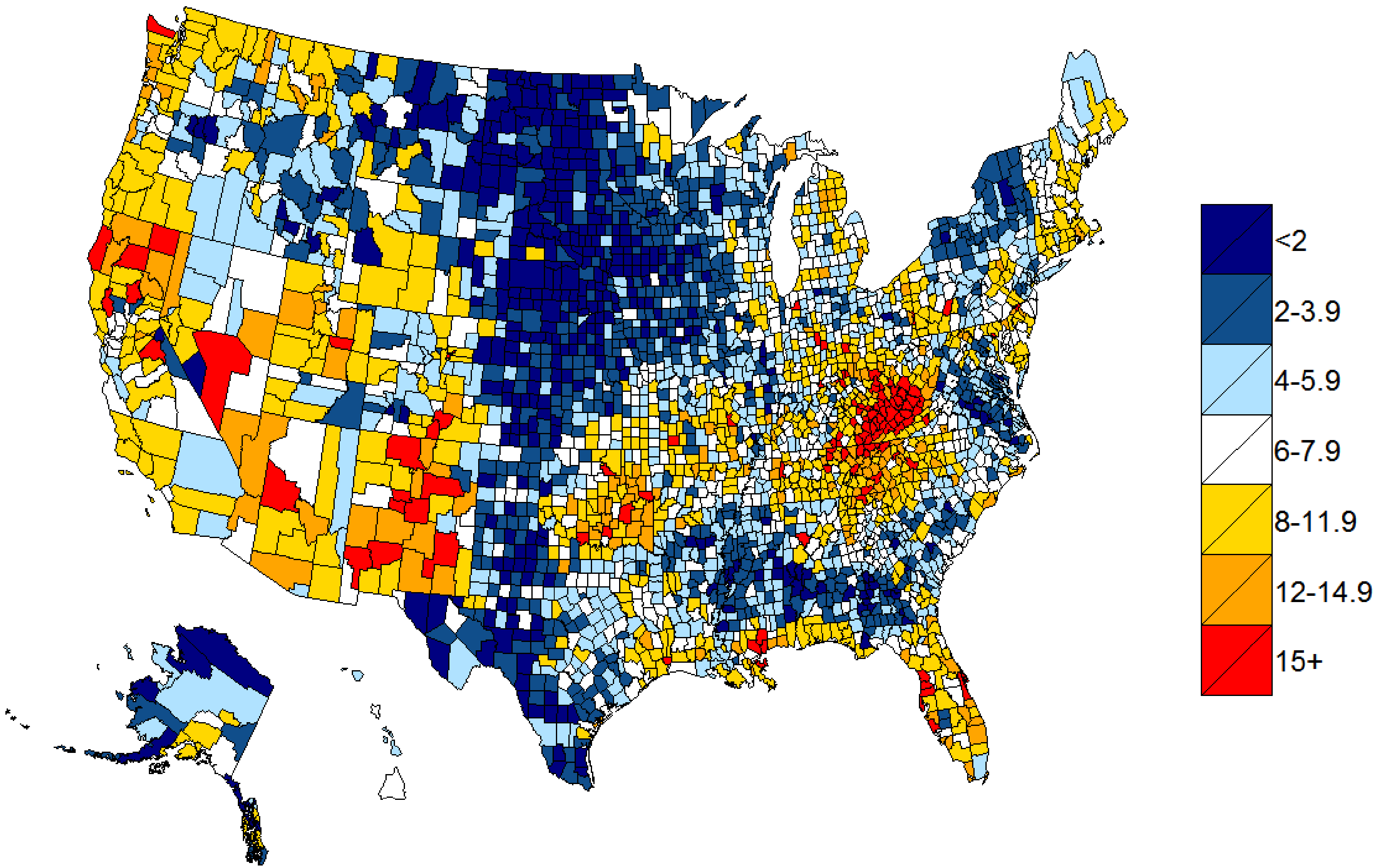
- Models run in Stata using GLAAMM (generalized linear latent and mixed models)

- Empirical Bayes predictions

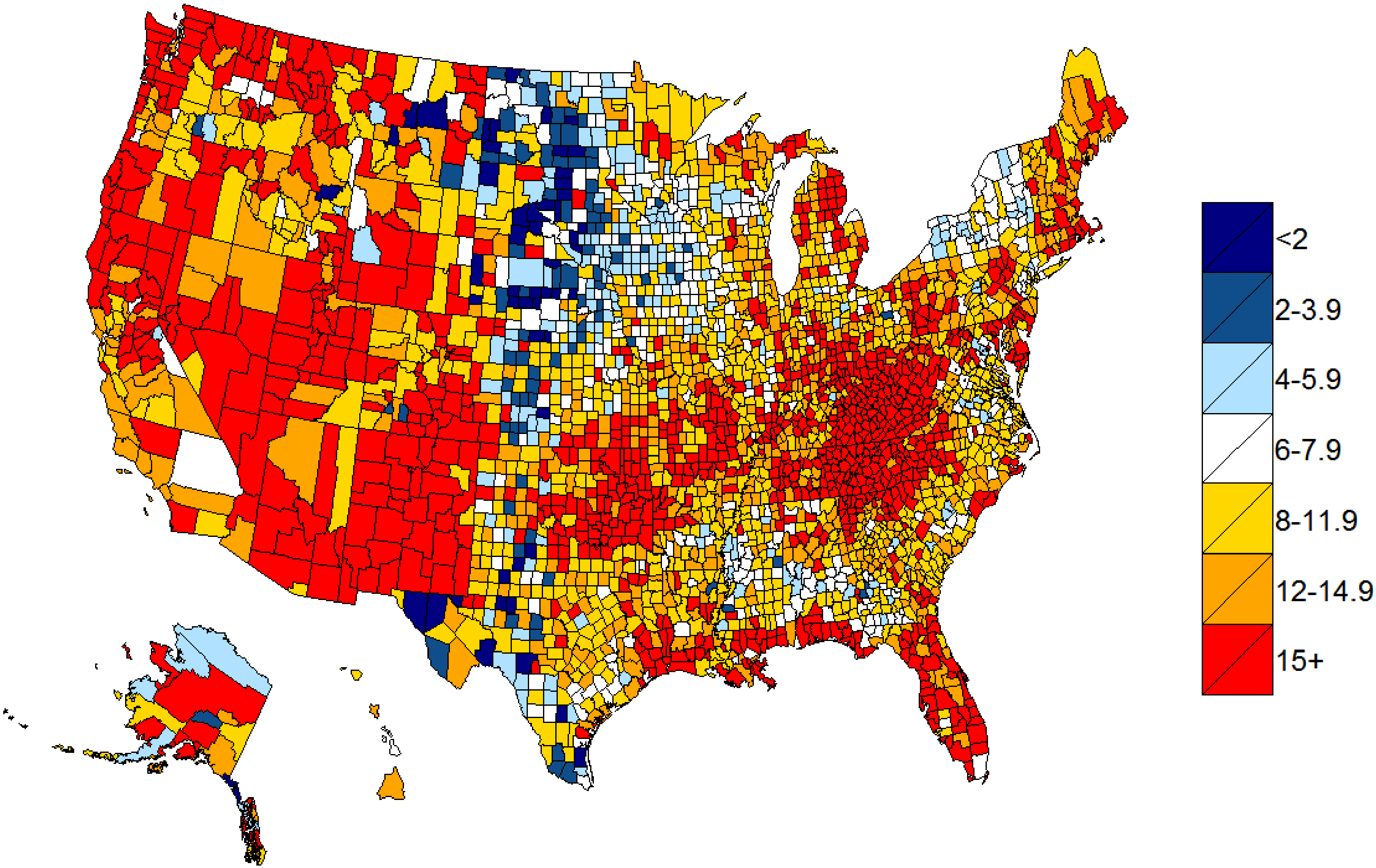
$$E(AADR) = [1\text{-}Pr(y_{it}=0)]*e^{\hat{y}_{it}}$$

- AADRs were mapped to examine spatiotemporal patterns
  - Hot and cold spots (Getis Ord Gi*)
    - Clusters of counties with high/low AADRs

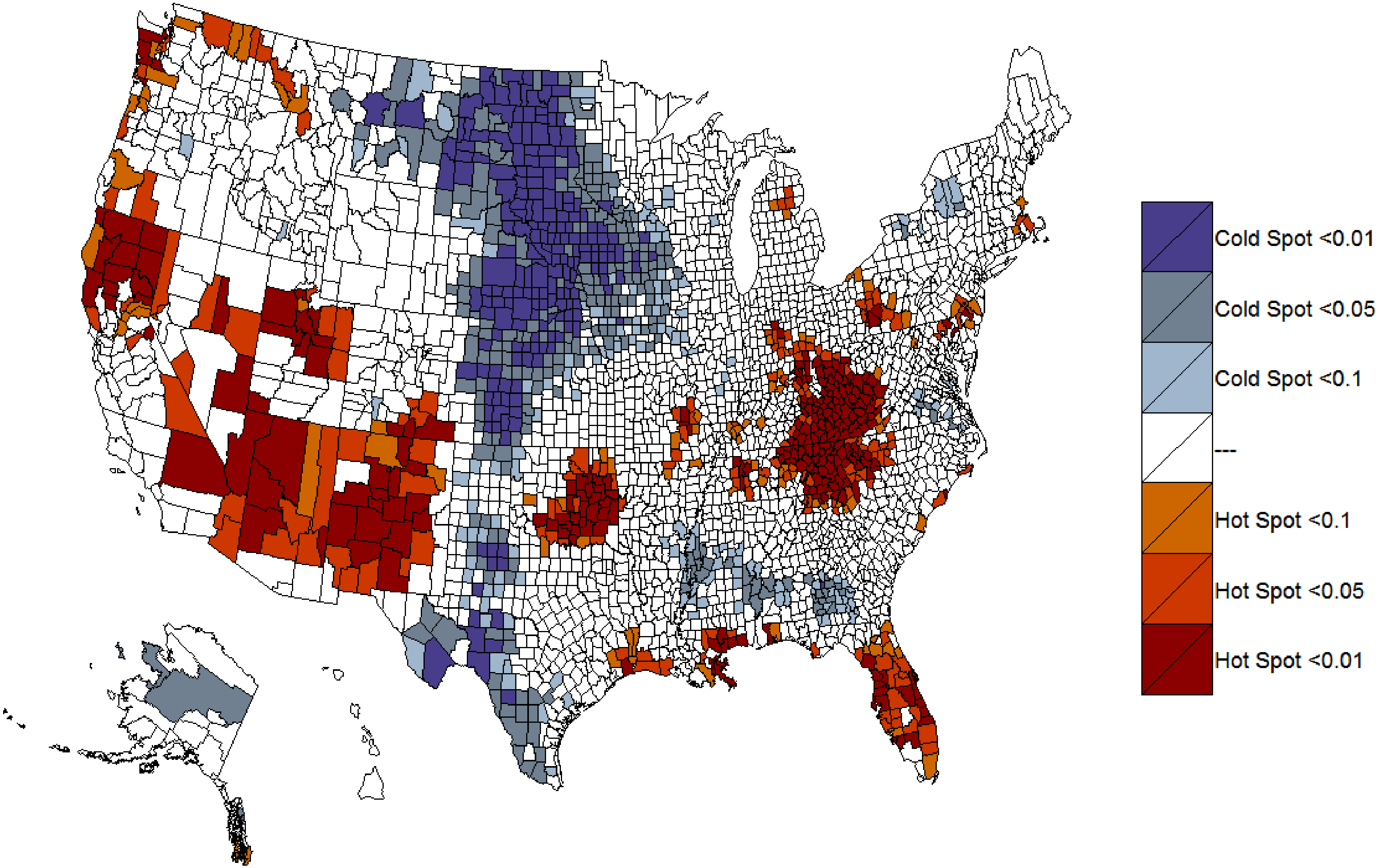
RESULTS: Age-adjusted death rates (per 100,000) due to drug poisoning - 2002



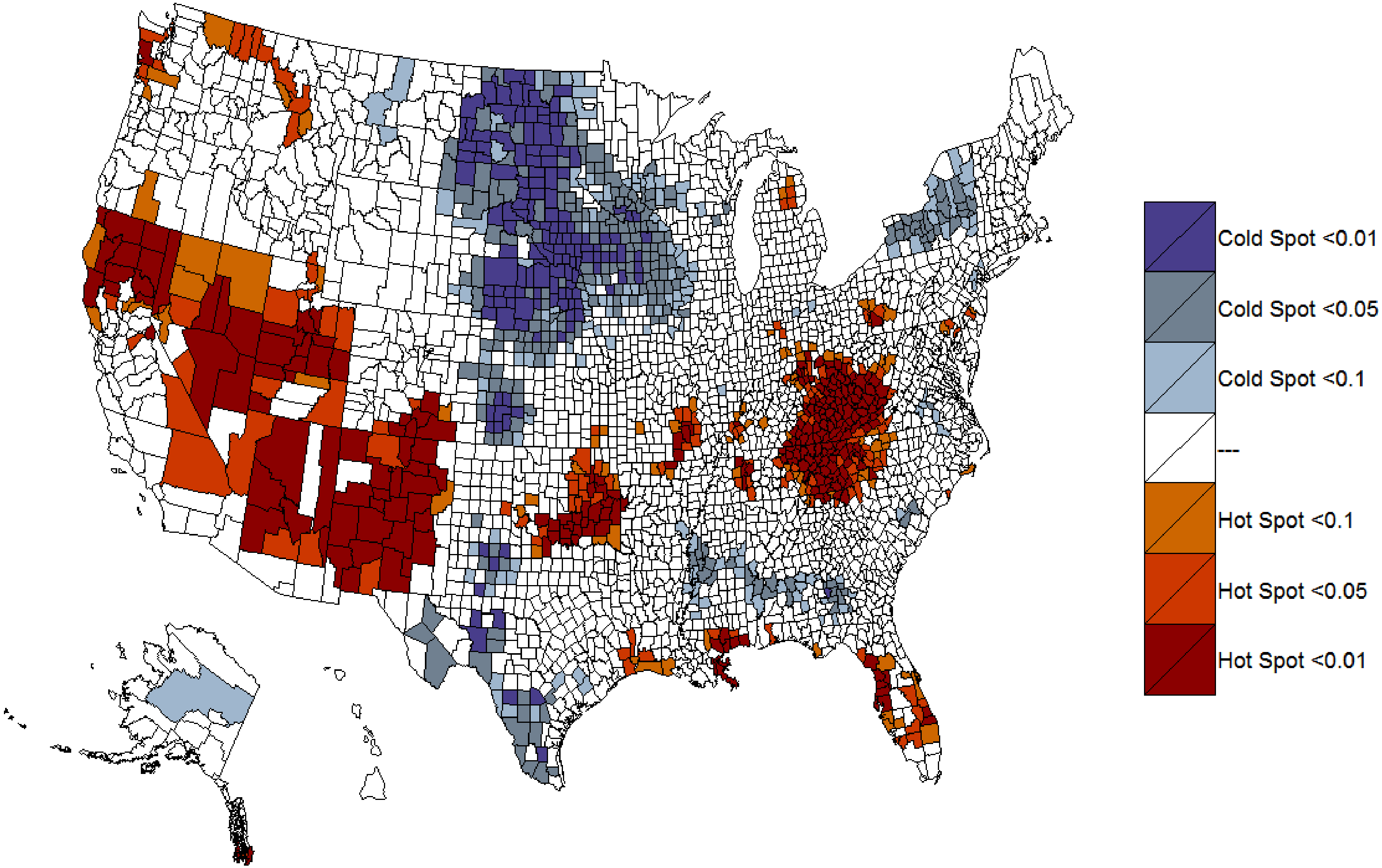
RESULTS: Age-adjusted death rates (per 100,000) due to drug poisoning - 2013



RESULTS: Hot and Cold Spots - 2002



RESULTS: Hot and Cold Spots - 2013



CONCLUSIONS

- Looking at spatiotemporal patterns can inform efforts to address drug poisoning mortality
  - Can help point to what might be driving drug poisoning mortality higher or lower in specific regions

- Patterns emerge that would have been missed using state estimates
  - Hot or cold spots that cross state boundaries
    - Appalachia, South West, Gulf coast
  - Significant sub-state variation
    - Mississippi, Montana, Virginia contain both hot and cold spots

Second Example

Teen Birth Rates in the U.S., 2003-2012

Teen Birth Rates in the U.S., 2003-2012

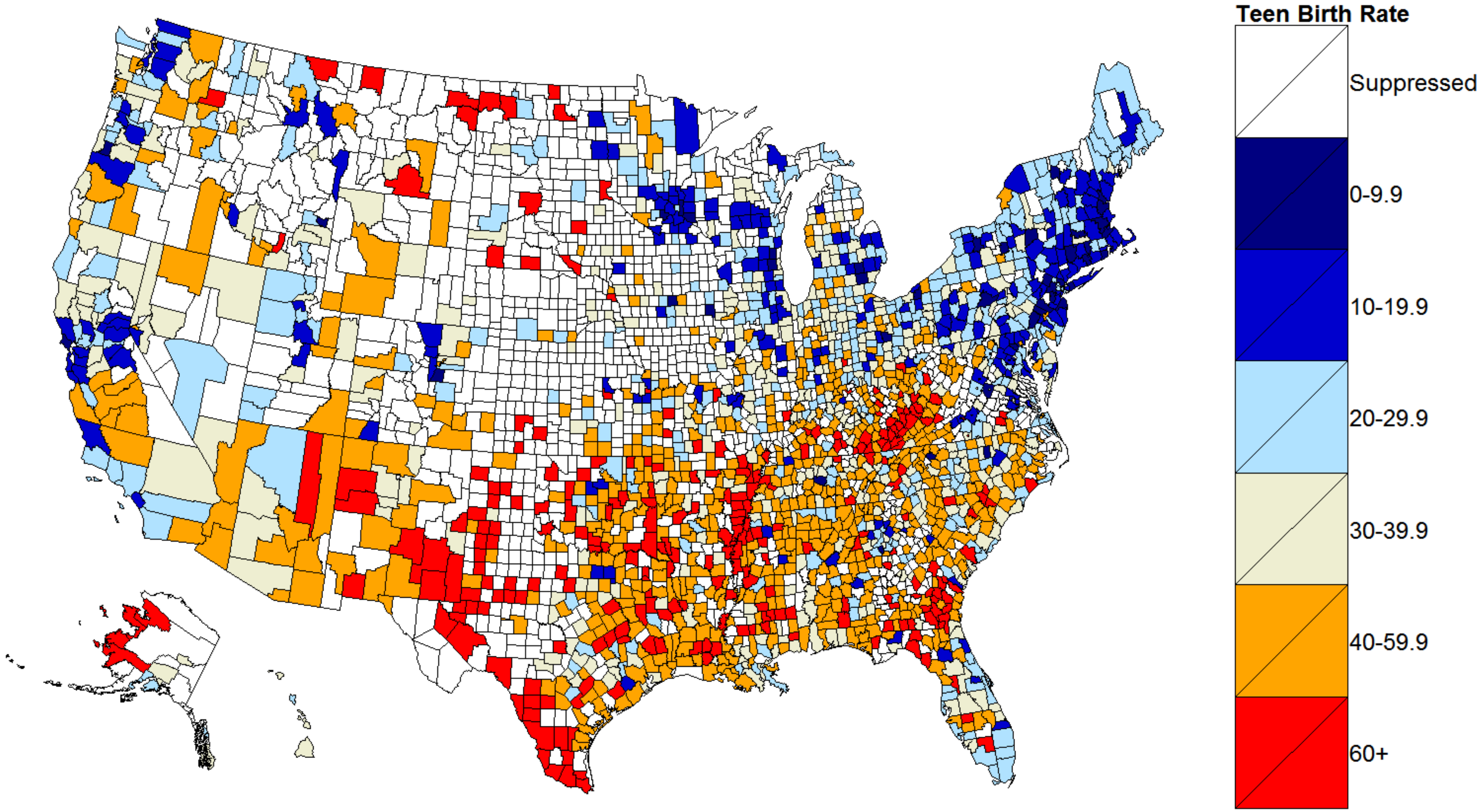
BACKGROUND

- In 2014, there were 24.2 births for every 1,000 adolescent females (15-19 years)

- Reducing teen pregnancy rates is a CDC Winnable Battle
  - Large-scale impact on health
  - Established preventive measures

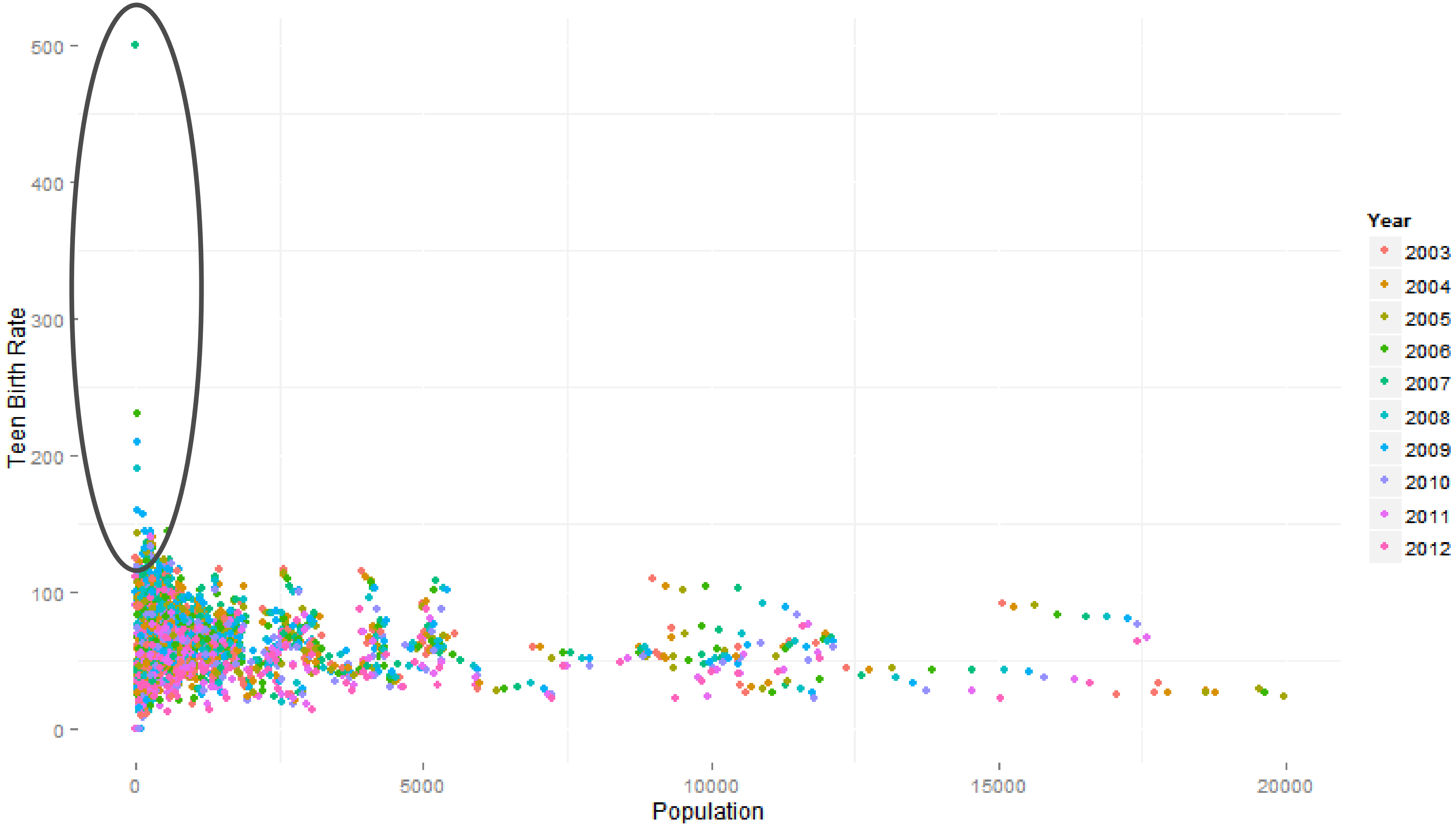
- Teen birth rates vary by state, as do trends over time
  - Spatiotemporal variation at the sub-state level has not yet been explored

RATIONALE FOR SMOOTHING: Teen Birth Rates



- Observed county-level teen birth rates in 2012
  - Suppressing counties with < 20 births (~36% counties)

RATIONALE FOR SMOOTHING (continued)



- Rates are unstable for counties with small populations (0 to **500** per 1,000)
- Could combine years, but that may mask temporal trends

DATA AND ANALYSES

y_{it} = # births to women 15-19 years of age in county *i* at time *t*

- National Vital Statistics Birth Data Files from 2003-2012

n_{it} = # women between 15-19 years in county *i* at time *t*

- bridged-race post-censal population estimates

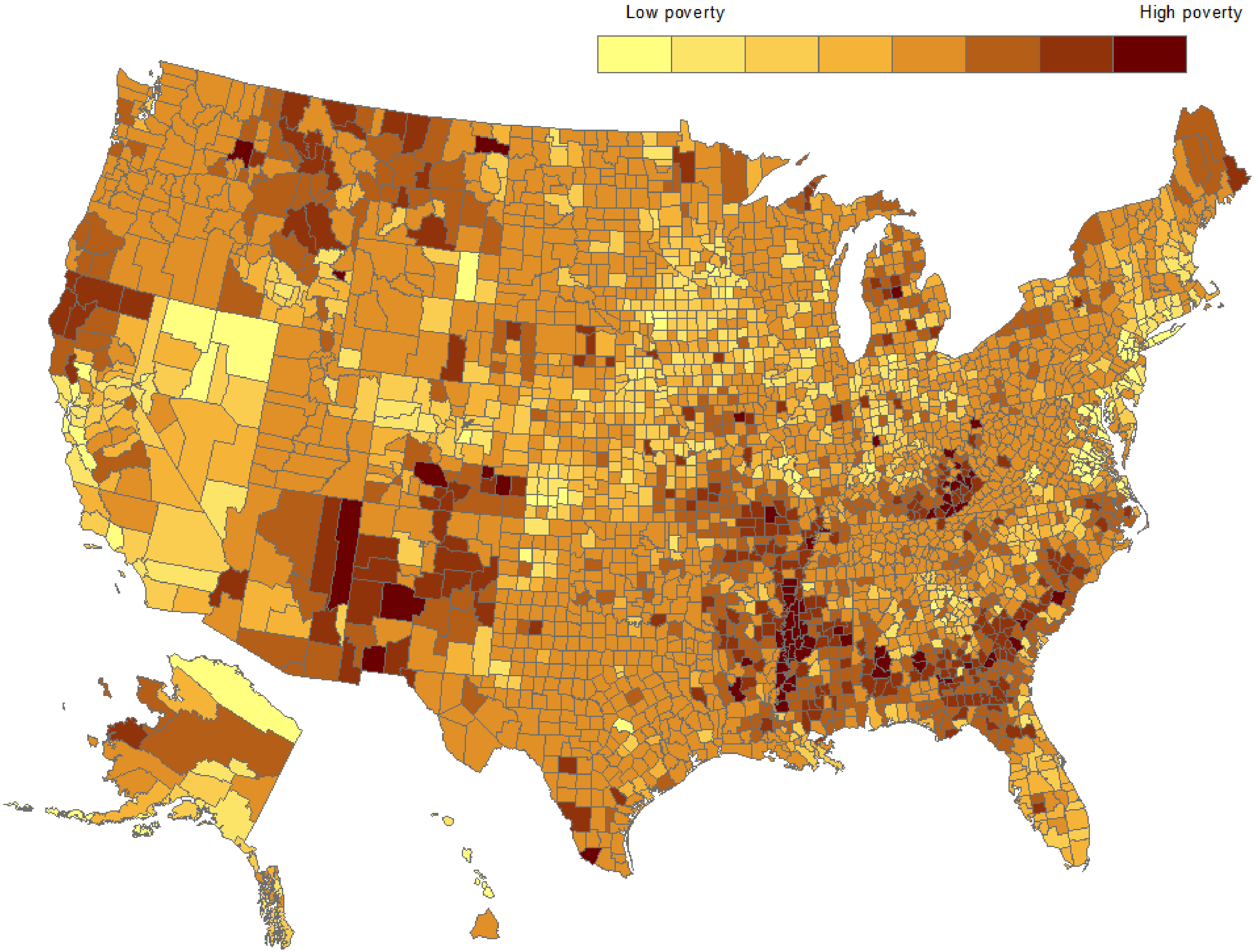
$y_{it} \sim$ Binomial(n_{it}, p_{it}), where,

p_{it} = the probabilities of teen birth for county *i* at time *t*

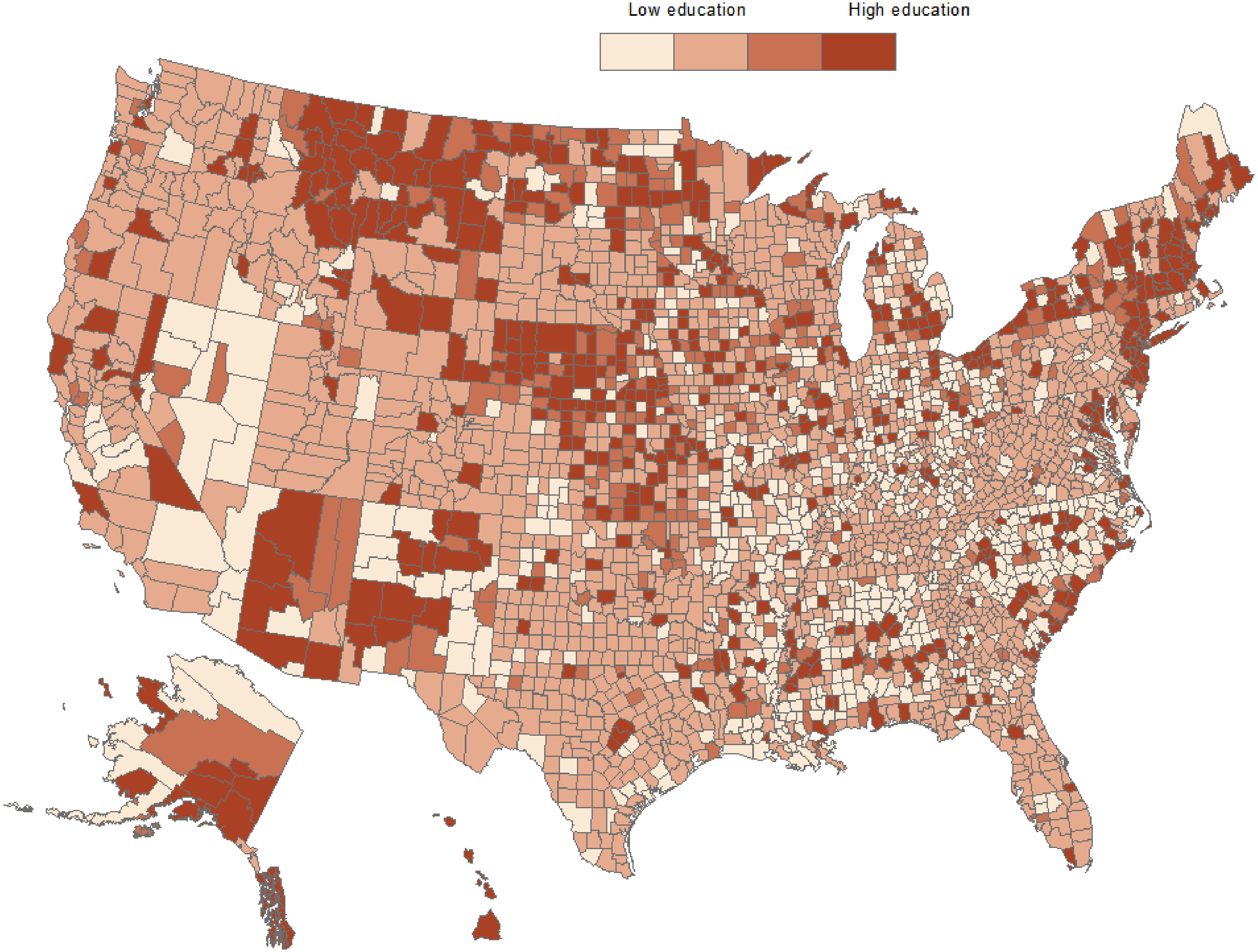
X_i' = set of covariates related to urban/rural designation, socio-demographic and economic characteristics

- Area Resource File, NCHS urban/rural classification

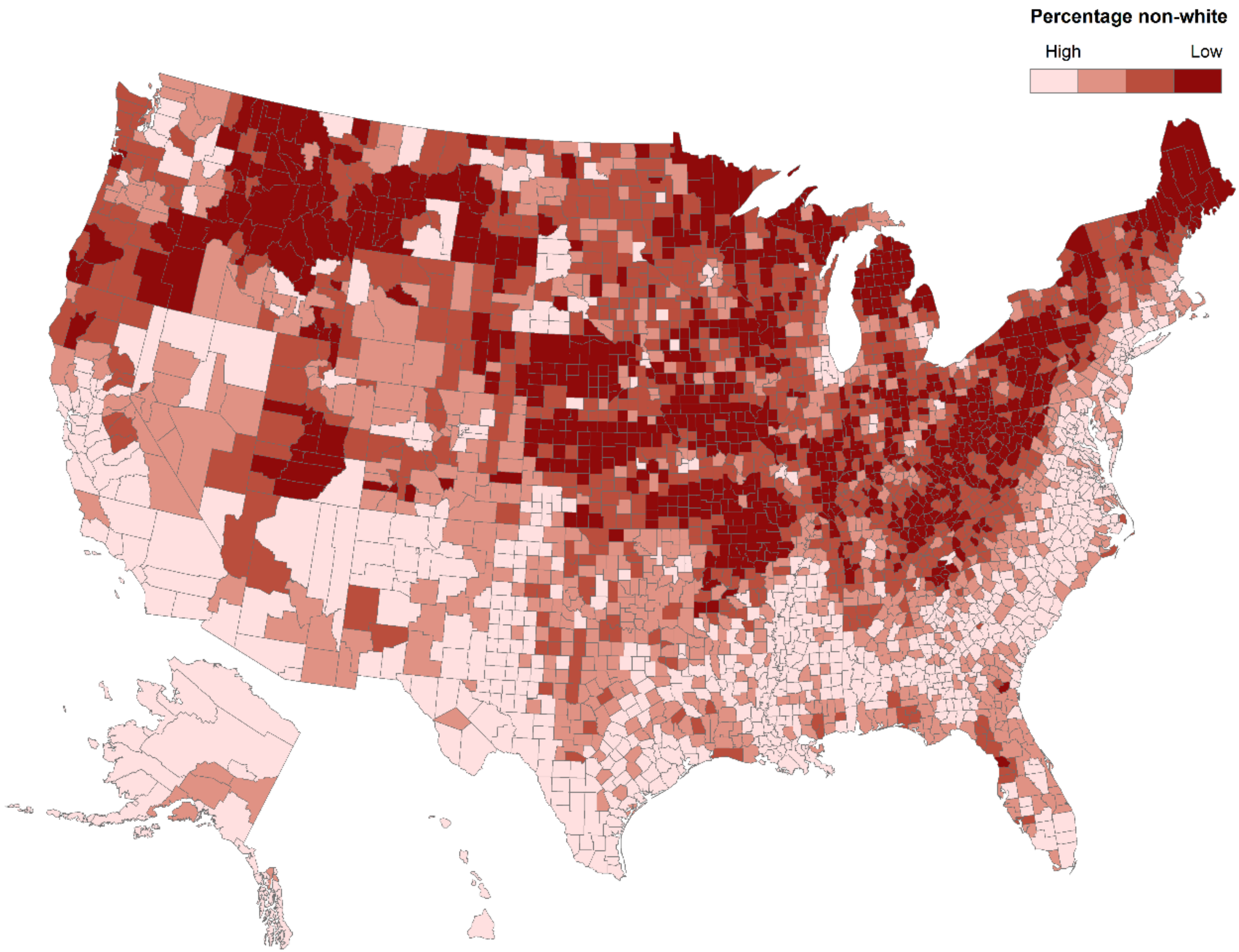
Covariates - Poverty



Covariates - Education



Covariates – Racial/Ethnic Distribution



HIERARCHICAL BAYESIAN MODELS

General space-time structure for modeling p_{it}:

$$\text{logit}(p_{it}) = \alpha + A_i + B_t + C_{it} + X_i'\gamma$$

α = intercept

A_i = spatial effect

B_t = temporal effect

C_{it} = space-time interaction

$X_i'\gamma$ = vector of covariates and corresponding parameters, γ

Models run in WinBUGS

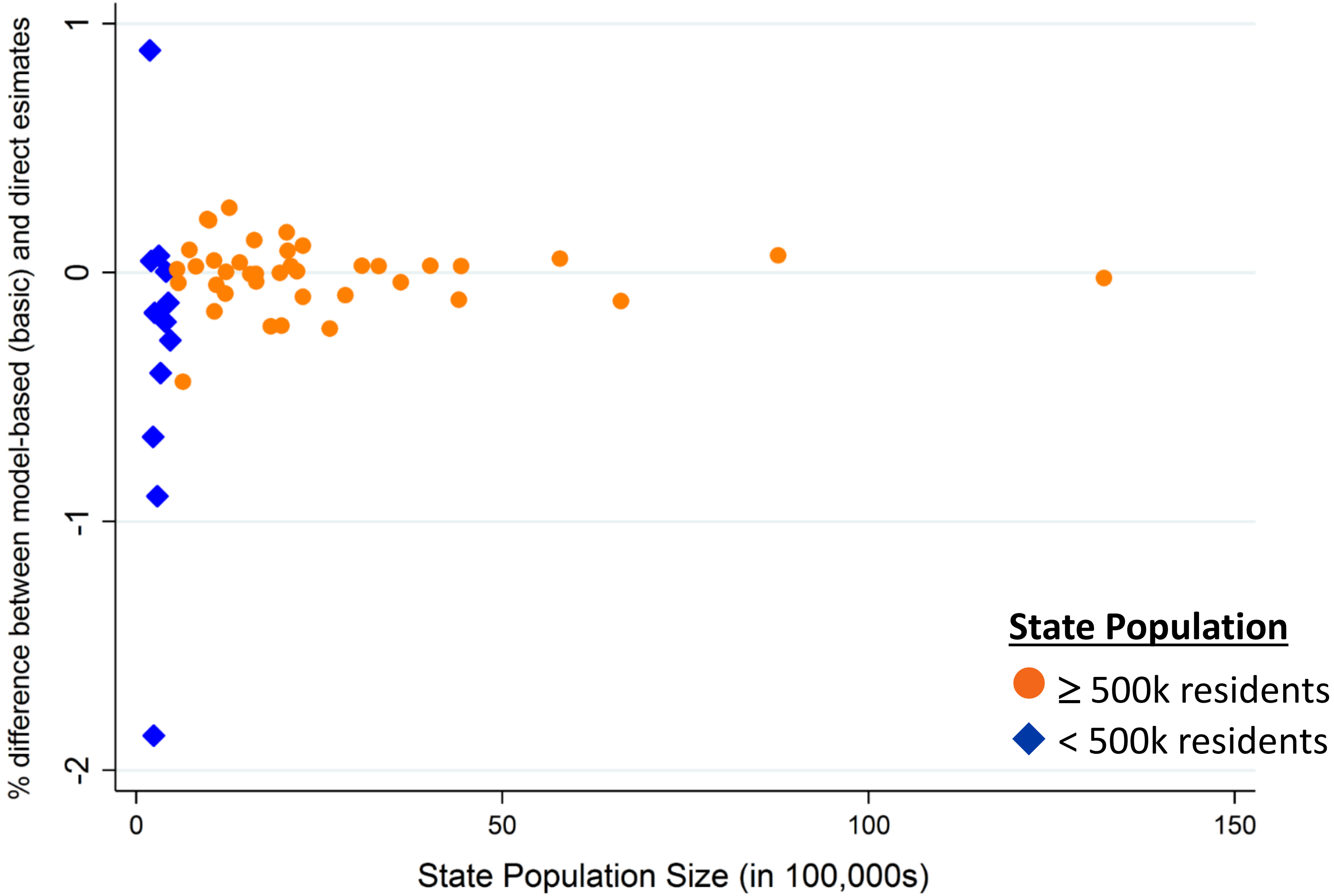
MAPPING SMOOTHED ESTIMATES

- Posterior teen birth rates ($1000*\hat{\hat{p}}_{it}$) mapped to examine spatiotemporal patterns:
  - Exceedance probabilities
    - Probability that counties exceed a specified threshold, *c*
      - *c* = 36 to reflect the mean county-level TBR in 2012
  - Hot and cold spots (Getis Ord Gi*)
    - Clusters of counties with high or low rates

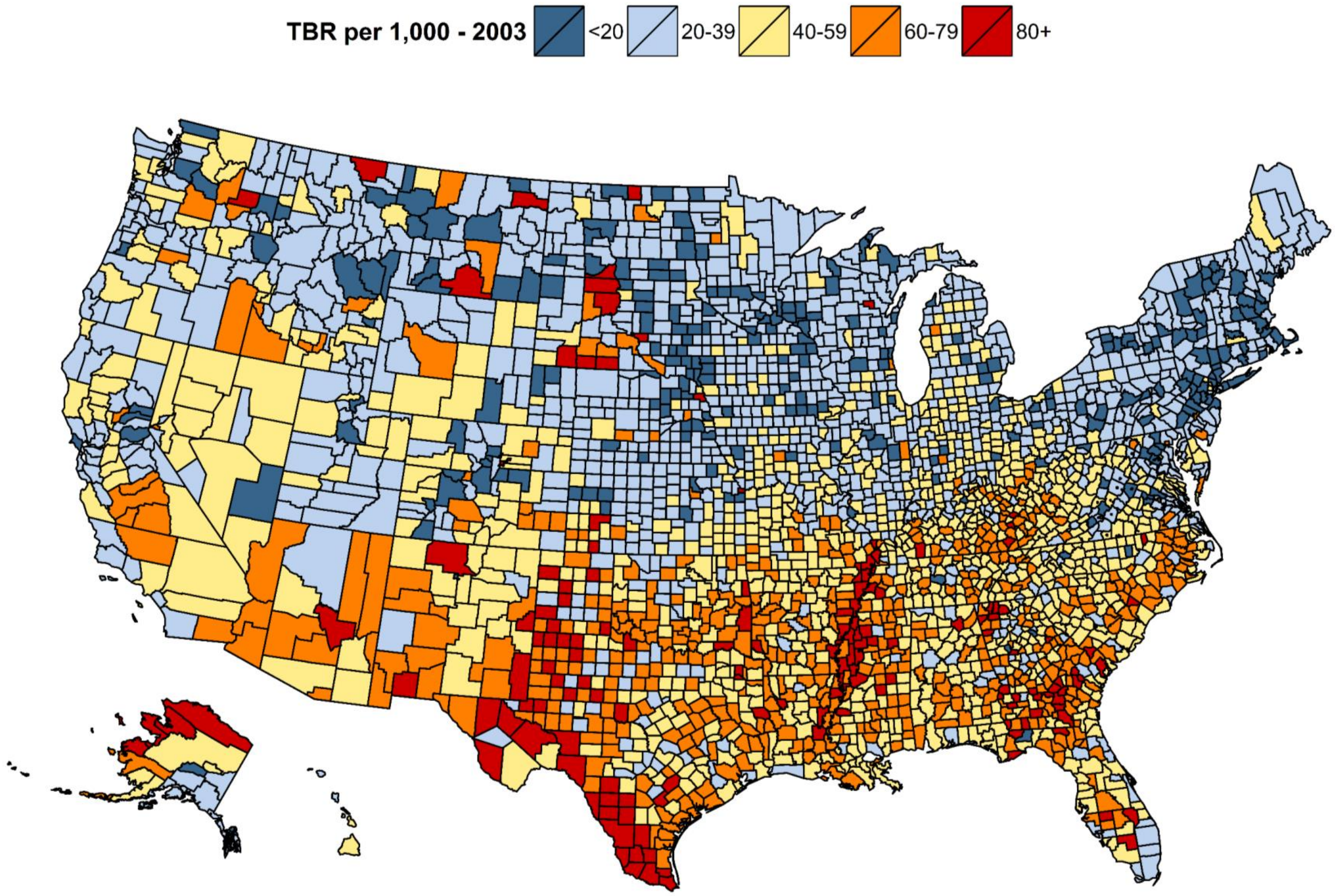
RESULTS

- From 2003-2012, teen birth rates:
  - declined for ~80% of counties
  - no change for ~19% of counties
  - increased for < 1% of counties
- Comparisons to direct estimates at the state level were within 2%
  - Differences between model-based and direct estimates were larger for sparsely populated states

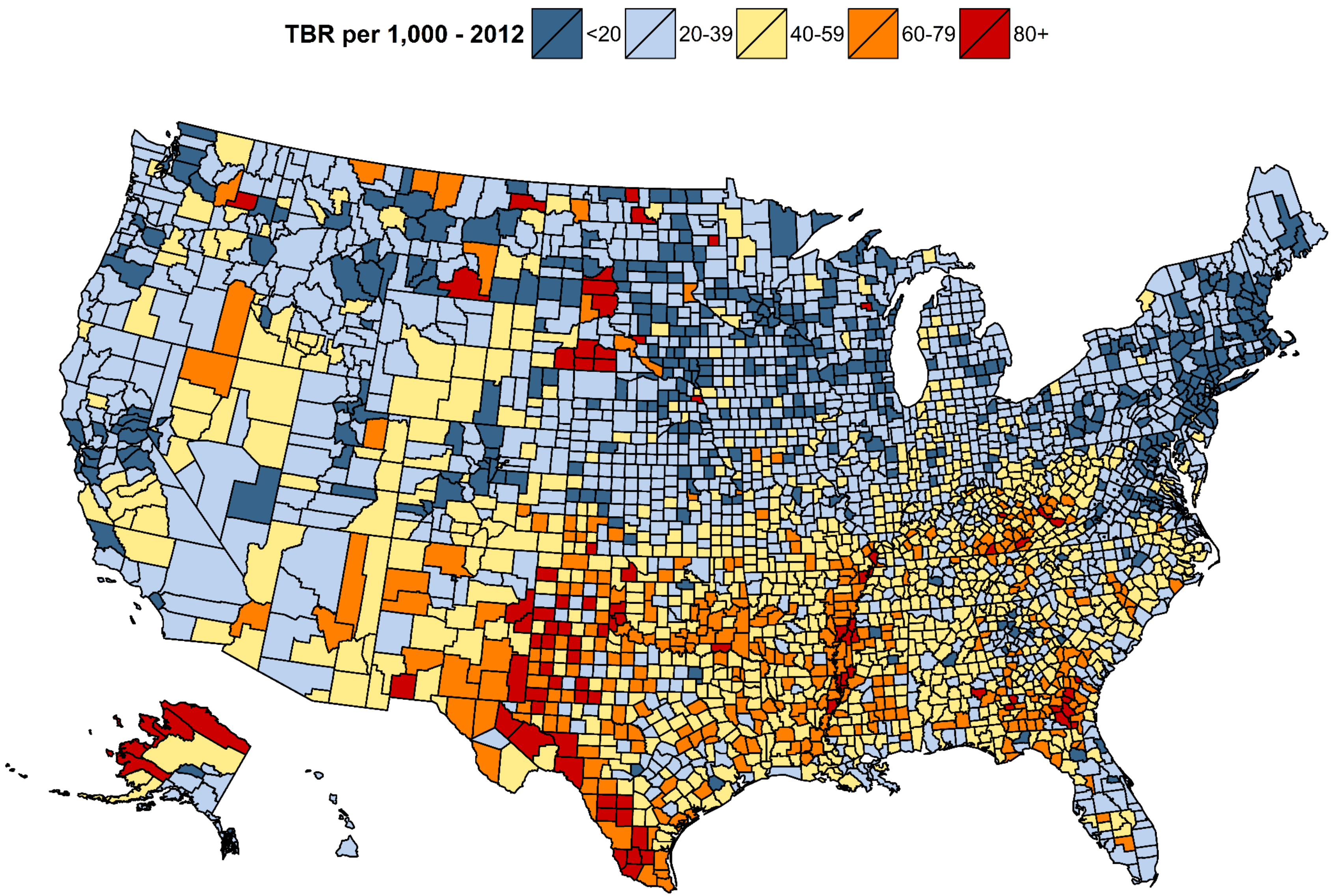
MODEL DIAGNOSTICS (Teen Birth Rates): Comparison to state estimates



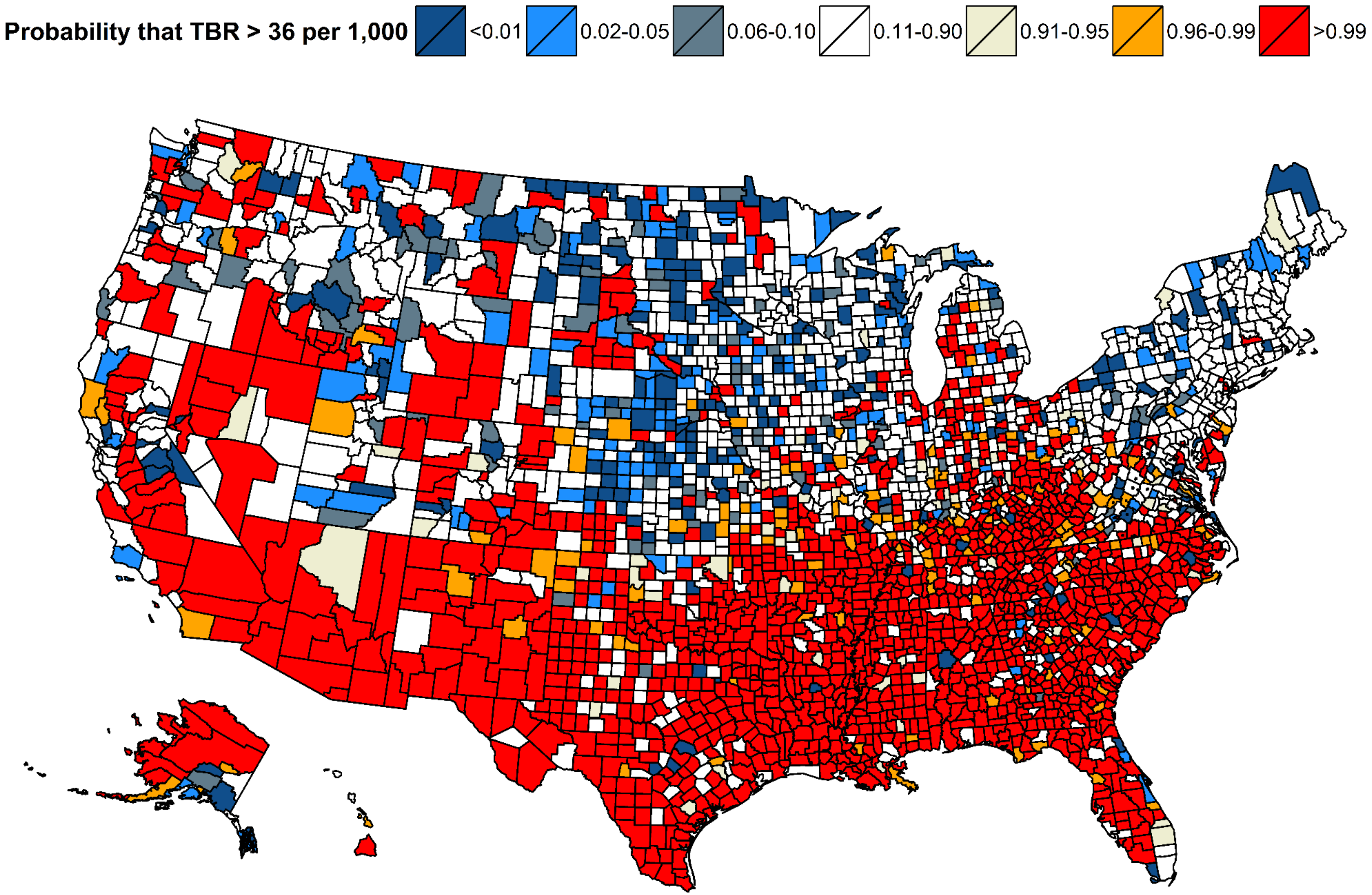
Smoothed teen birth rates (per 1,000) - 2003



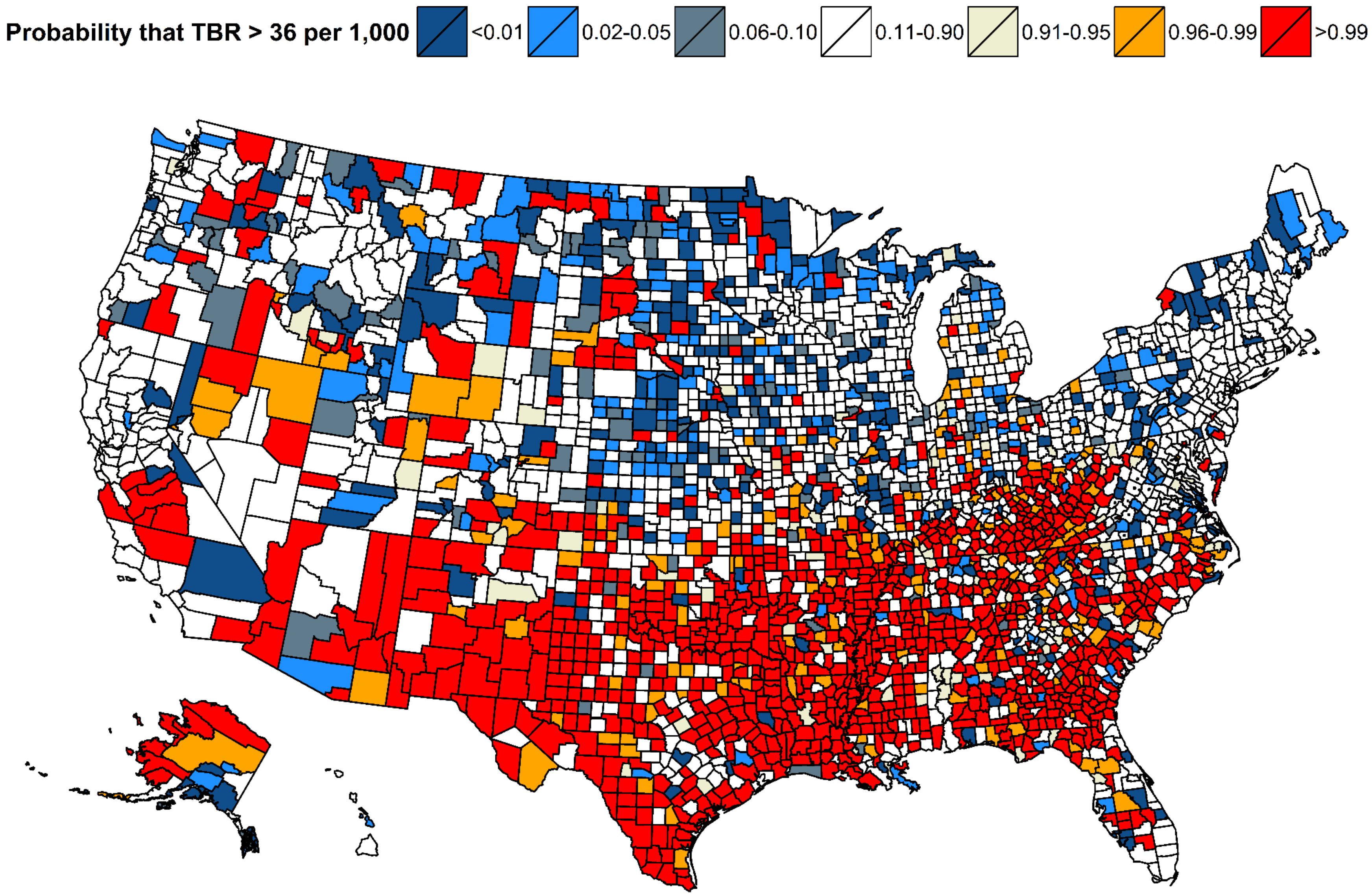
Smoothed teen birth rates (per 1,000) - 2012



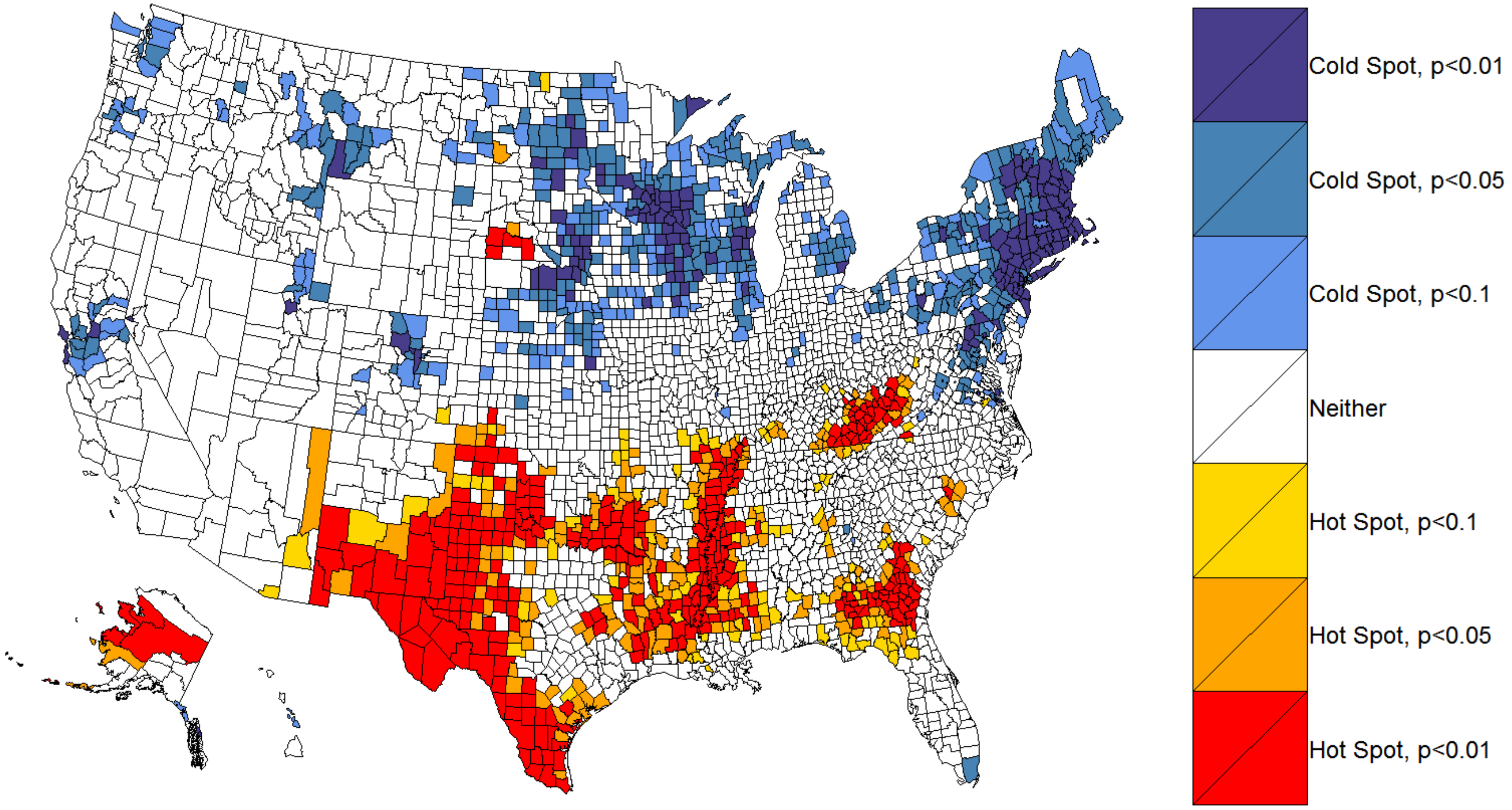
Exceedance Probabilities - 2003



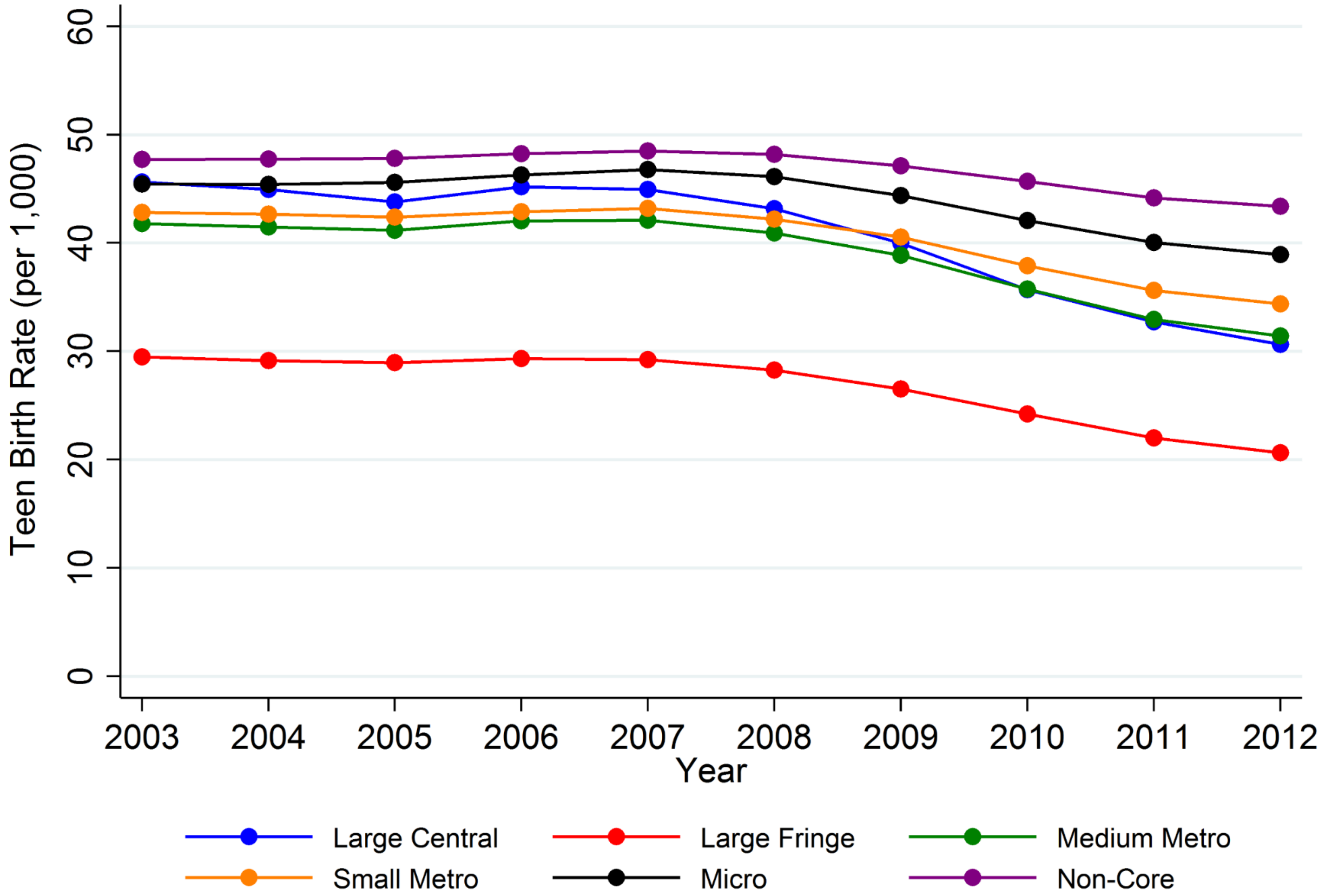
Exceedance Probabilities - 2012



Hot and Cold Spots - 2012



Trends by Urban/Rural Designation, Teen Birth Rates 2003-2012



CONCLUSIONS

- Findings highlight counties where teen birth rates are relatively higher or lower
  - How trends over time vary geographically
- Patterns emerge that we would have missed using state estimates
  - For example, the hot spot along the Mississippi River crosses state boundaries
- Examination of spatiotemporal patterns may inform efforts to further reduce birth rates to adolescents in the U.S.
  - Can look at where teen birth rates are higher than a given ‘target’

SOME CONSIDERATIONS

- Strengths and opportunities:
  - Can see and examine variation across the U.S.
  - Pick up on important patterns that might be masked by state estimates or other groupings (urban/rural)
  - Provide information relevant to public health efforts at the state or local level
  - Shed light on risk/protective factors associated with population health outcomes

SOME CONSIDERATIONS

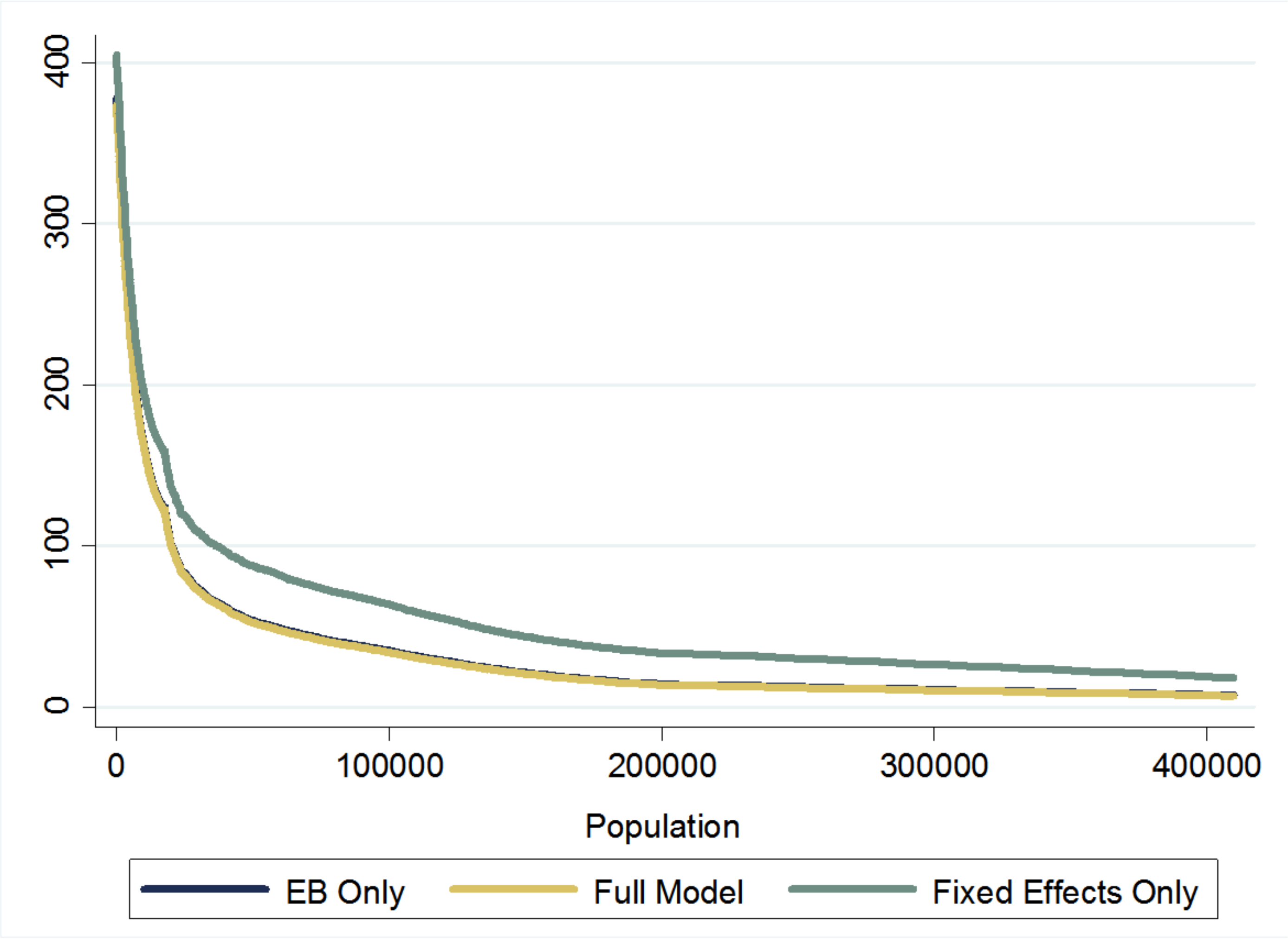
- Limitations and challenges:
  - Model-based estimates might smooth away important effects (either in space or time)
  - Some analyses are VERY computer intensive
    - 6+ weeks running on a 32 GB machine
  - Might not have the level of geography we want
    - Is county the appropriate unit of geography?
  - Data are typically restricted-use
    - Implications for access, confidentiality

QUESTIONS?

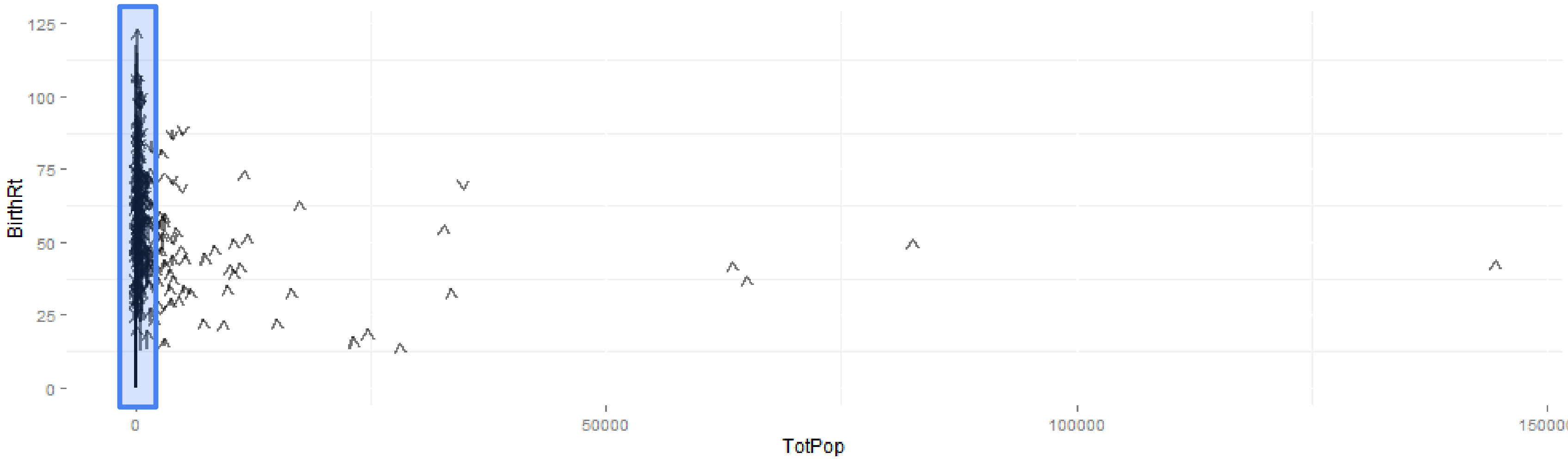
Email: LRossen@cdc.gov

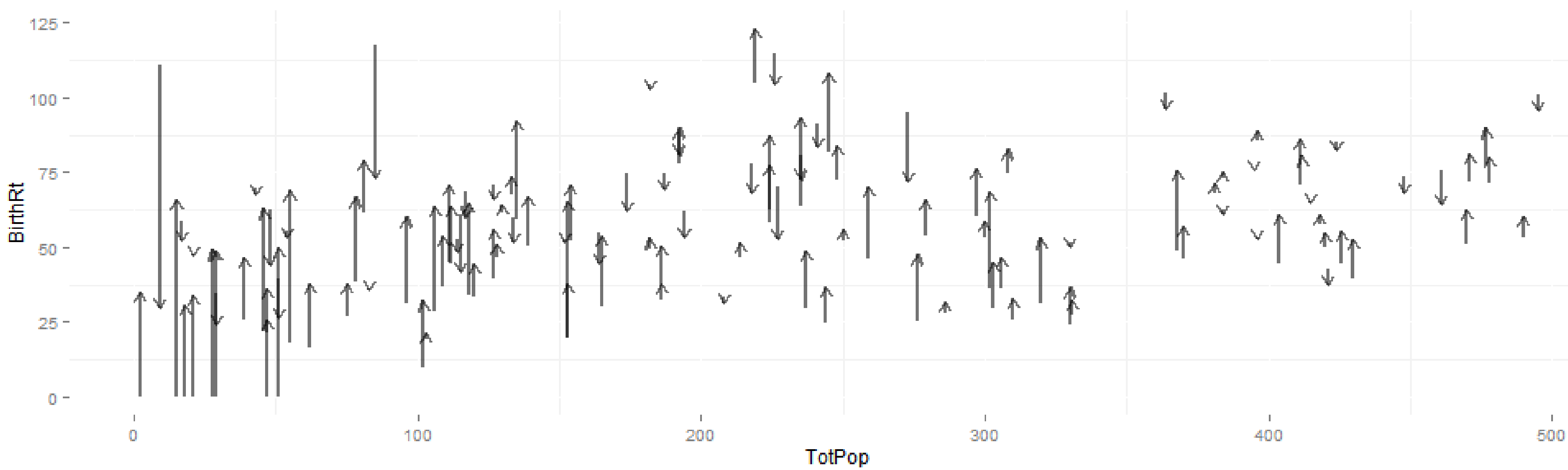
EXTRA SLIDES

MODEL DIAGNOSTICS (Drug Poisoning): $(Y_{obs} - Y_{pred})^2$ vs. Population Size



MODEL DIAGNOSTICS (Teen Birth): Effects of shrinkage





Helpful References

NCHS Fact Sheet: Data on Drug Poisoning Deaths. June 2015. http://www.cdc.gov/nchs/data/factsheets/factsheet_drug_poisoning.pdf

Tiwari C, Beyer K, Rushton G. The impact of data suppression on local mortality rates: The case of CDC WONDER. Am J Public Health. 2014;104(8):1386-1388. doi: 10.2105/AJPH.2014.301900

Rossen LM, Khan D, Warner M. Trends and geographic patterns in drug-poisoning death rates in the U.S., 1999-2009. Am J Prev Med. 2013;45(6):e19-25. doi: 10.1016/j.amepre.2013.07.012.

Rossen LM, Khan D, Warner M. Hot spots in mortality from drug poisoning in the United States, 2007-2009. Health Place. 2014;26:14-20. doi: 10.1016/j.healthplace.2013.11.005

Helpful References

Skrondal A, Rabe-Hesketh S. Prediction in multilevel generalized linear models. J Royal Statistical Society: Series A (Statistics in Society). 2009;172: 659–687. doi: 10.1111/j.1467-985X.2009.00587.x

Hamilton BE, Martin JA, Osterman MJK, Curtin SC. Births: Preliminary Data for 2014. National Vital Statistics Reports. Volume 64, Number 6. http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_06.pdf

Carlin BP, Louis TA. 2009. Bayesian Methods for Data Analysis. New York: Chapman and Hall.

Lawson A. 2013. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. New York: Chapman and Hall.