Flexible Models for Elucidating Health Disparities

by Eric C. Tassone, J.D., Ph.D., Alan E. Gelfand, Ph.D., & Marie Lynn Miranda, Ph.D.

Children's Environmental Health Initiative

Children's Environmental Health Initiative

- Focus on children
  - Focus on issues of environmental justice
    - Shift to preventive interventions
      - Emphasis on spatial analytic approaches

Outline of Talk

- Overview: “What”, “how”, and “why” of disaggregation
- Modeling Details
- Computing Details and Issues
- Example: North Carolina Detailed Birth Record Data
- Impact on Policy?
- Future Directions

Overview

- **Disaggregation? What, how, and why…**
  - **What:**
    - **A method that substantially extends inferential possibilities of customary modeling of spatial outcomes data that are areally observed**
    - **Computationally tractable for large data sets (on the order of 10^5 or more)**
    - **'Disaggregated': we disaggregate aggregated counts from usual spatial model into subgroups using individual-level characteristics**

Overview

- **Disaggregation? What, how, and why…**
  - **How:**
    - **Model subgroups in areal unit using individual-level data**
    - **Multi-way contingency table for each areal unit**
    - **Explained with loglinear model in each areal unit**
    - **Spatially smooth models via random effects**
    - **All in a multilevel modeling framework**

Overview

- **Disaggregation? What, how, and why…**
  - **Why?:**
    - **Other approaches problematic with subgroups**
      - **Sometimes adjust via covariates or expecteds**
      - **How to model the subgroups? ANOVA-like approach? Ind. models? Via multivariate CAR priors? Ignore / aggregate?**
      - **Confined to ‘outcome’ cond. on ‘risk factors’**
  - **Use available individual-level data..true level?**

Overview

- **Disaggregation? What, how, and why… more why…**
  - **Flexible inference in multilevel structure**
  - **Dimension reduction (r << L in general)**
  - **No need to specify a "response" variable**
  - **Joint modeling=>arbitrary marginal and cond. probs.**
    - **Not just conditional probability statements**
    - **Arbitrary marginal, joint, and cond. statements**
    - **Flexible aggregation: investigate outcomes/ groups of interest, e.g., racial disparities**

Modeling Details

Cell counts:

$$n_l^{(s)} \sim Po(\lambda_l^{(s)})$$

First level:

$$\log(\lambda_l^{(s)}) = \mathbf{X}_l^T \beta_{\mathbf{s}} + \log(n_\cdot^{(s)}) = \sum_{t=1}^{r} X_{lt}\beta_{st} + \log(n_\cdot^{(s)})$$

Second level:

$$\beta_{st} = \mathbf{w}_s^T \eta_t + \tilde{\phi}_t^{(s)} = \sum_{u=1}^{q} w_{su}\eta_{tu} + \tilde{\phi}_t^{(s)}$$

Random effects:

$$\tilde{\phi}_t^{(s)} = \phi_t^0 + \phi_t^{(s)}$$

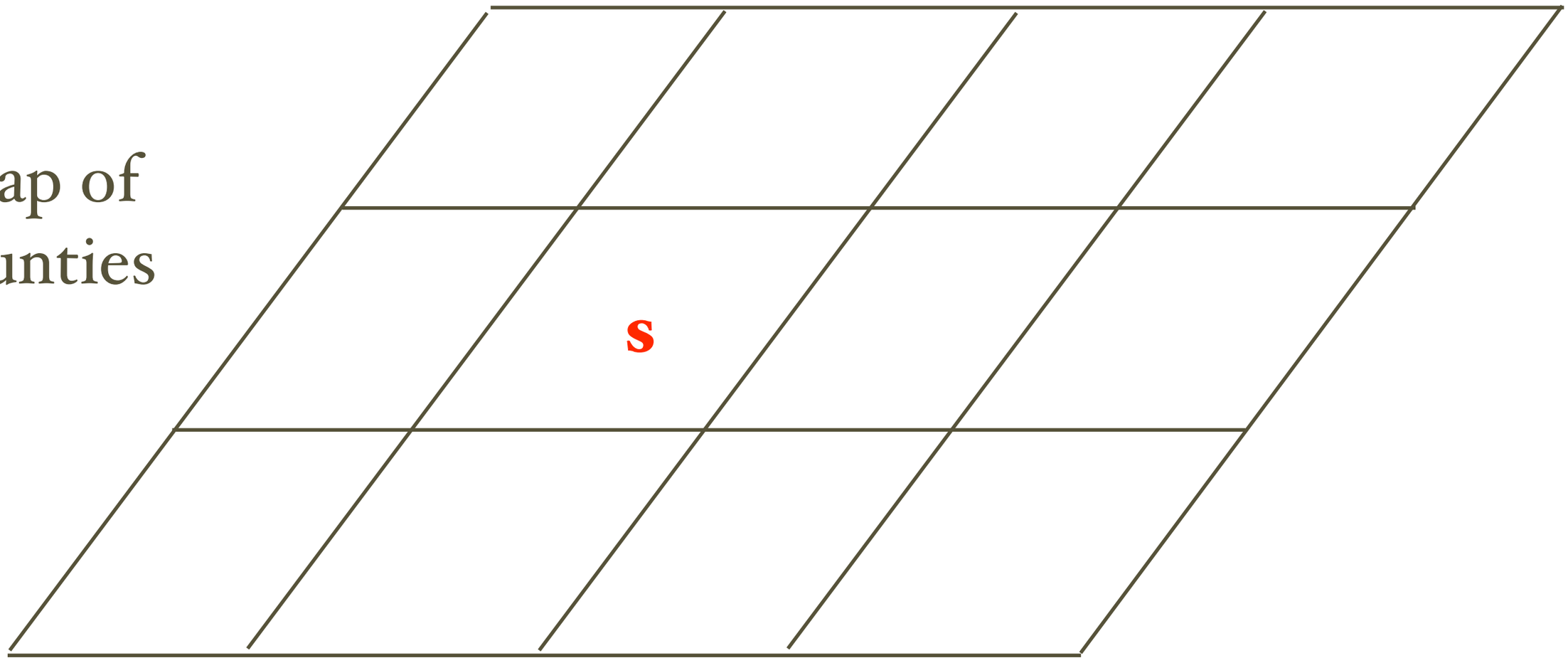
And we can plug the second level into the first level for very nice interpretations of the overall model, namely the log counts as...

$$\mathbf{X}\boldsymbol{\phi}^0 + \mathbf{X}\mathbf{W}_s\boldsymbol{\eta} + \mathbf{X}\boldsymbol{\phi}^{(s)} \text{ .}$$

(+ offset)

Spatial Smoothing -- The phi_s's

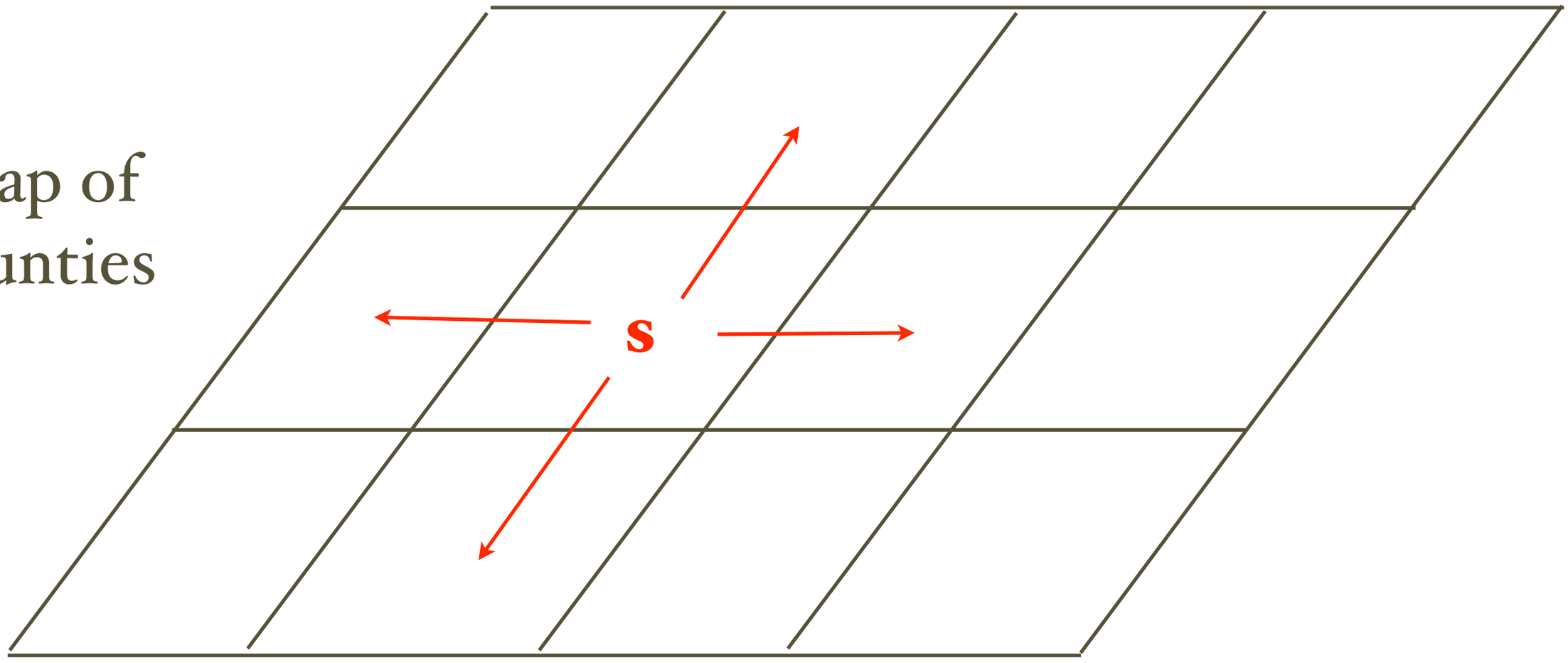
Map of counties



Smoothes county **s**'s beta_st to be like its neighbors (for selected t's).

Spatial Smoothing -- The phi_s's

Map of counties



Smoothes county **s**'s beta_st to be like its neighbors (for selected t's).

Computing Details

- Can be implemented in WinBUGS ... computationally OK
- Many modeling choices in this flexible framework...
  - Design matrix for loglinear model...
  - Which loglin. model parameters get spatial smooth?
  - Which get areal unit-level covariates? Which covars.?
  - What form for spatial random effects?
    - Independent?
    - Multivariate CAR?
    - Attempt some sort of dimension reduction?
- For the example we'll see, we will detail our choices

Example

- NC Detailed Birth Record
  - 1999-2003
  - No congenital anomalies
    - Singletons
    - N=463,639, with 32,437 LBW (~6.996%)
- County-level (though finer resolution available)
  - ZCTA … alternative state analysis?
  - Census Tract, Block Group, Block…focused analysis?

Example

- **Variables**
  - **X: Maternal race (African Amer. (AA) or white)**
  - **Y: Low birth weight (yes or no)**
  - **Z: Sex of infant (female or male)**
  - **W: Maternal tobacco use (yes or no) - "Smoking"**
- **So 2 x 2 x 2 x 2 = 16 subgroups in each areal unit**

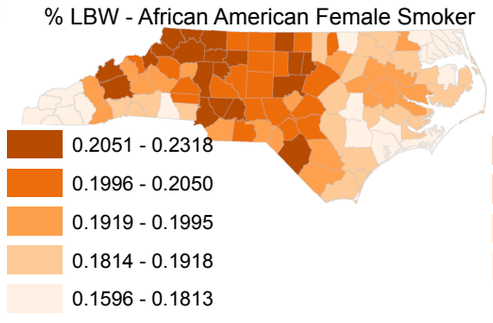
Example

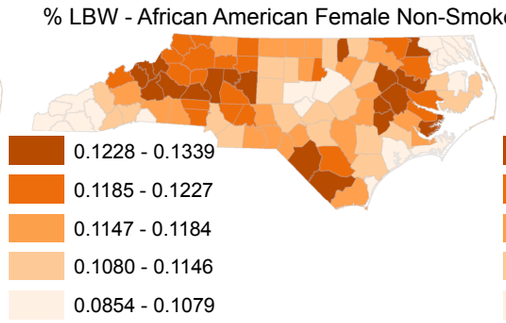
- Let L denote subgroup, l=1, 2, …, 16
- Let S=1, 2, …, 100 index counties in NC
- Model provides $p_l^{(s)}$, the probability for subgroup l in county s … not prob. LBW given X,Z,W, but joint prob..
  - Re-combine into cond., marginal, etc., probs.
  - Disparity measures (e.g., odds ratio)
- Model fit: (XYZ, XYW, YZW)
  - So reduce from 16 to 14 … more reduction in general

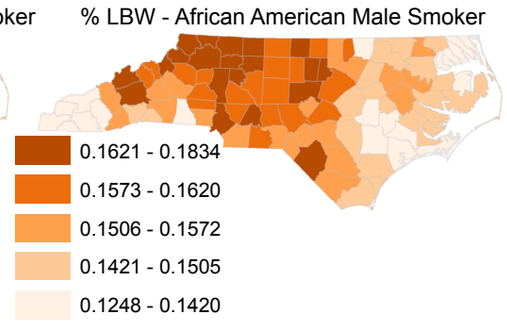
Example: Model

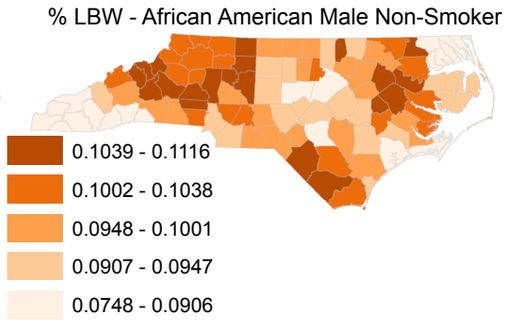
$$\log \lambda_{ijkm}^{(s)} = \gamma^{(s)}$$
$$+ \gamma_i^{X(s)} + \gamma_j^{Y(s)} + \gamma_k^{Z} + \gamma_m^{W(s)}$$
$$+ \gamma_{ij}^{XY(s)} + \gamma_{ik}^{XZ(s)} + \gamma_{im}^{XW(s)} \gamma_{jk}^{YZ(s)} + \gamma_{jm}^{YW(s)} + \gamma_{km}^{ZW(s)}$$
$$+ \gamma_{ijk}^{XYZ} + \gamma_{ijm}^{XYW} + \gamma_{jkm}^{YZW} + \log(n_{\cdot}^{(s)}).$$

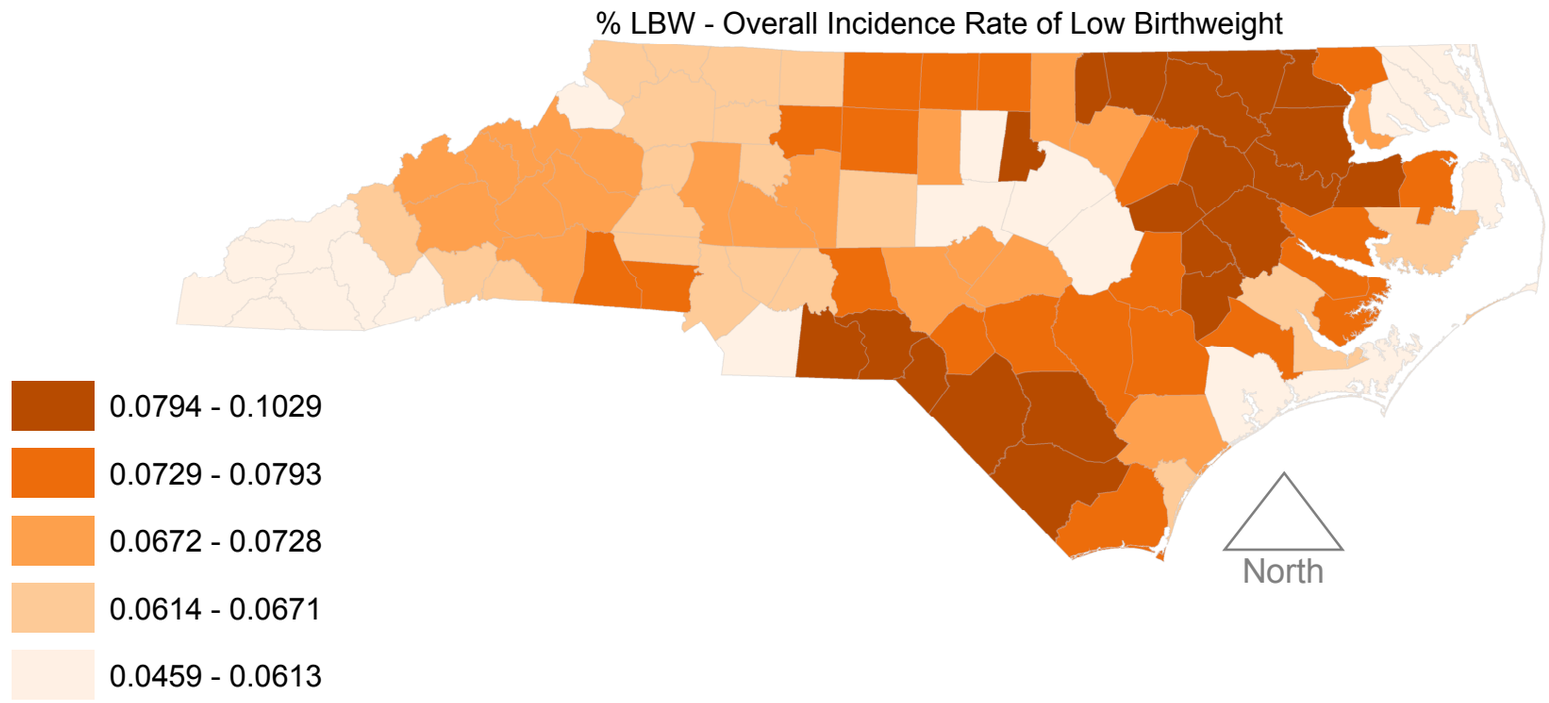
Est. LBW%, Overall (center) and Subgroups

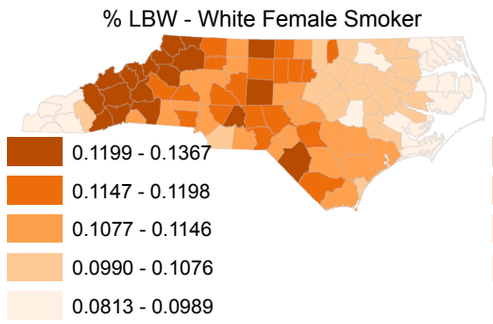


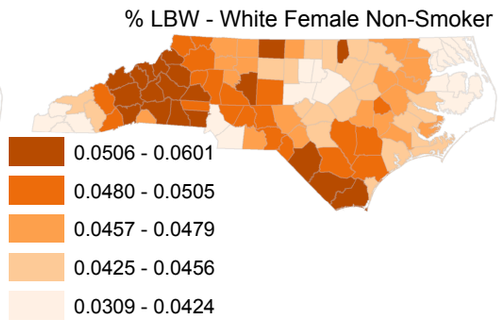


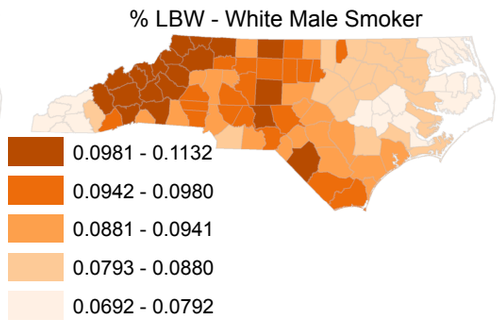


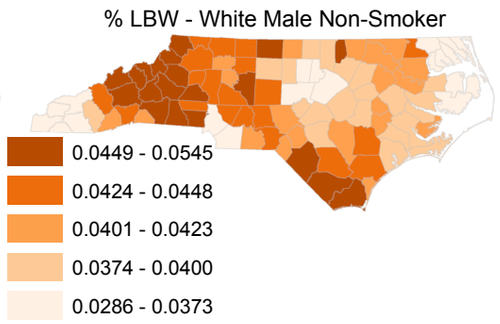






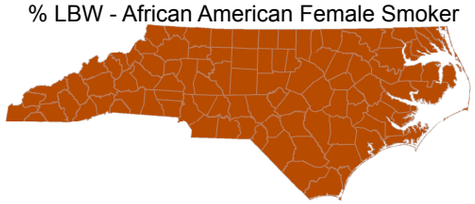




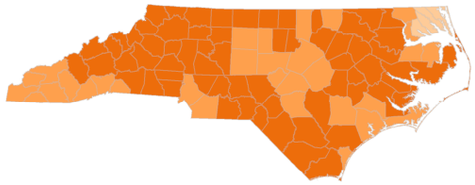


Est. LBW%, Overall (center) and Subgroups

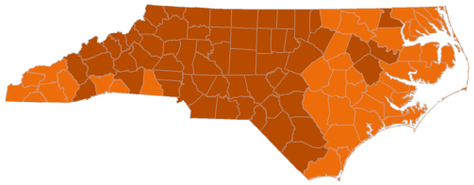
% LBW - African American Female Smoker



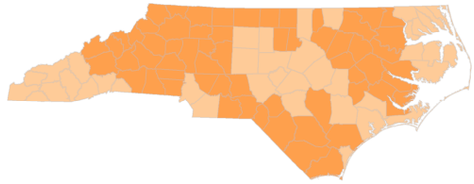
% LBW - African American Female Non-Smoker



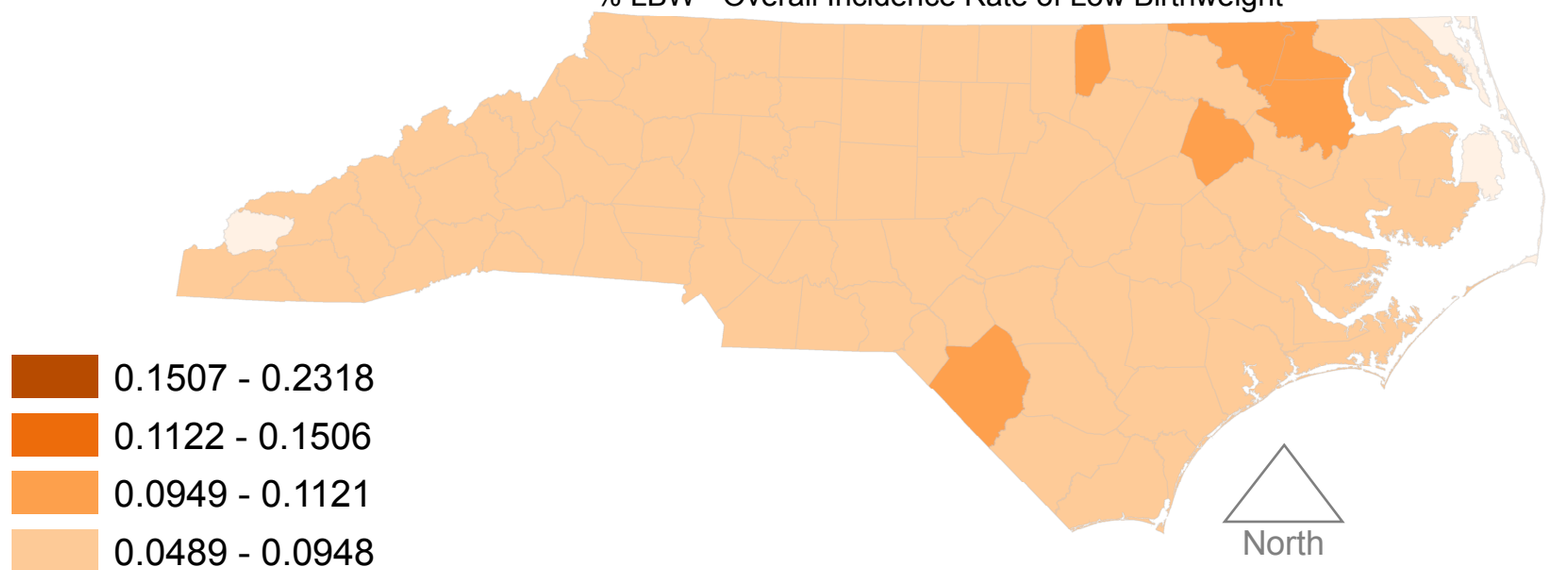
% LBW - African American Male Smoker



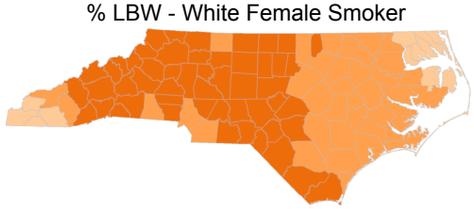
% LBW - African American Male Non-Smoker



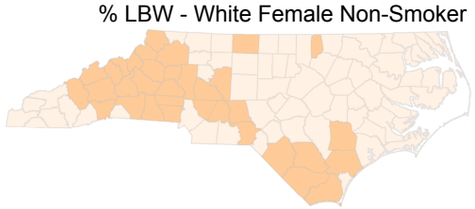
% LBW - Overall Incidence Rate of Low Birthweight



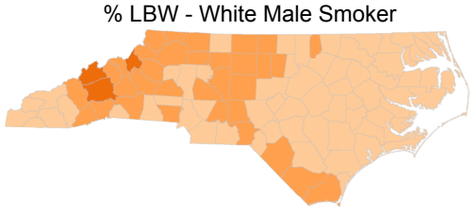
% LBW - White Female Smoker



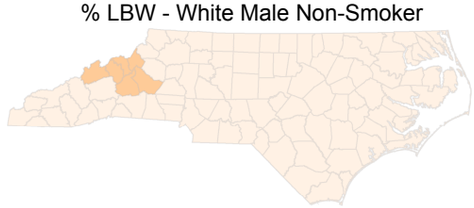
% LBW - White Female Non-Smoker



% LBW - White Male Smoker

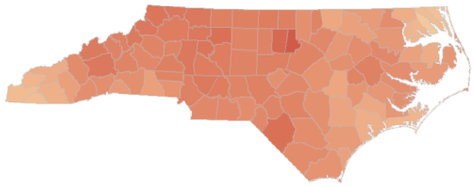


% LBW - White Male Non-Smoker

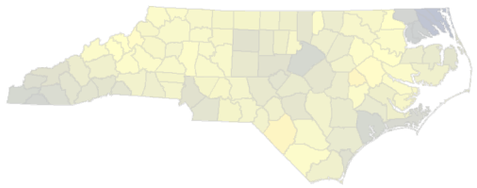


Est. LBW%, Overall (center) and Subgroups

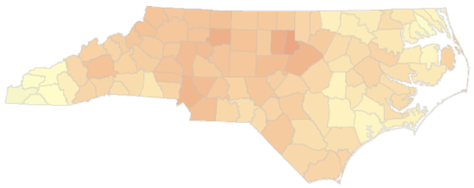
Girl, AA Smoker



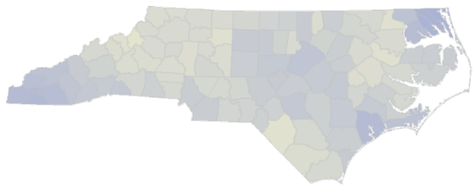
Girl, AA nonSm



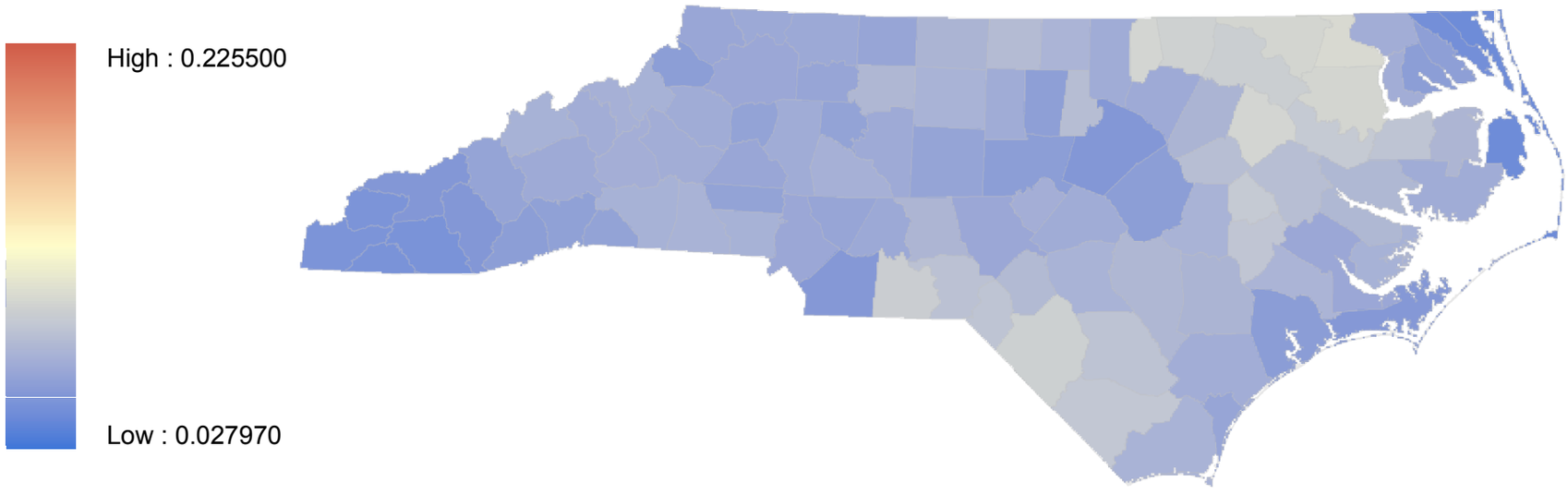
Boy, AA Smoker



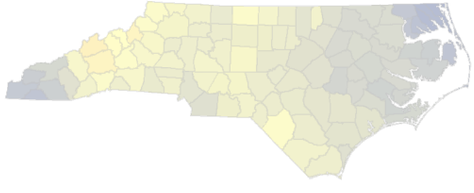
Boy, AA nonSm



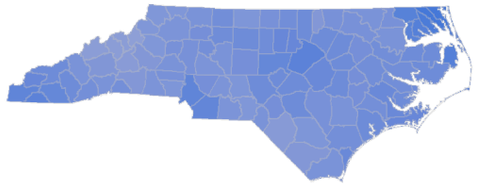
All births



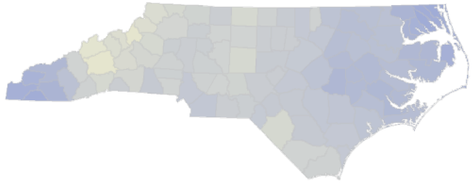
Girl, W Smoker



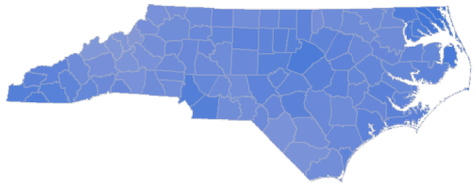
Girl, W nonSm



Boy, W Smoker

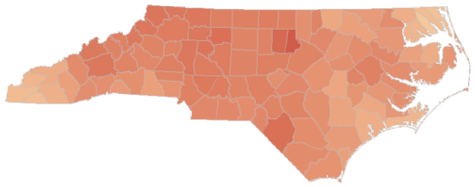


Boy, W nonSm

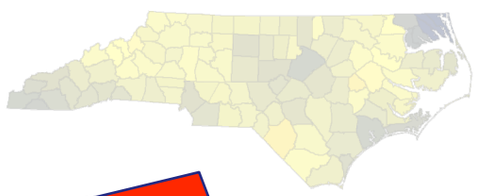


Est. LBW%, Overall (center) and Subgroups

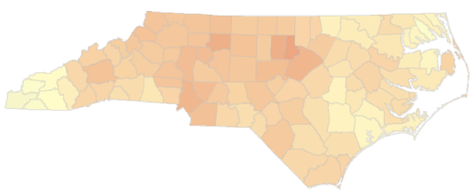
Girl, AA Smoker



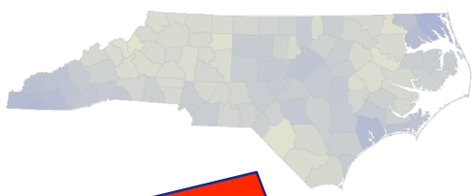
Girl, AA nonSm



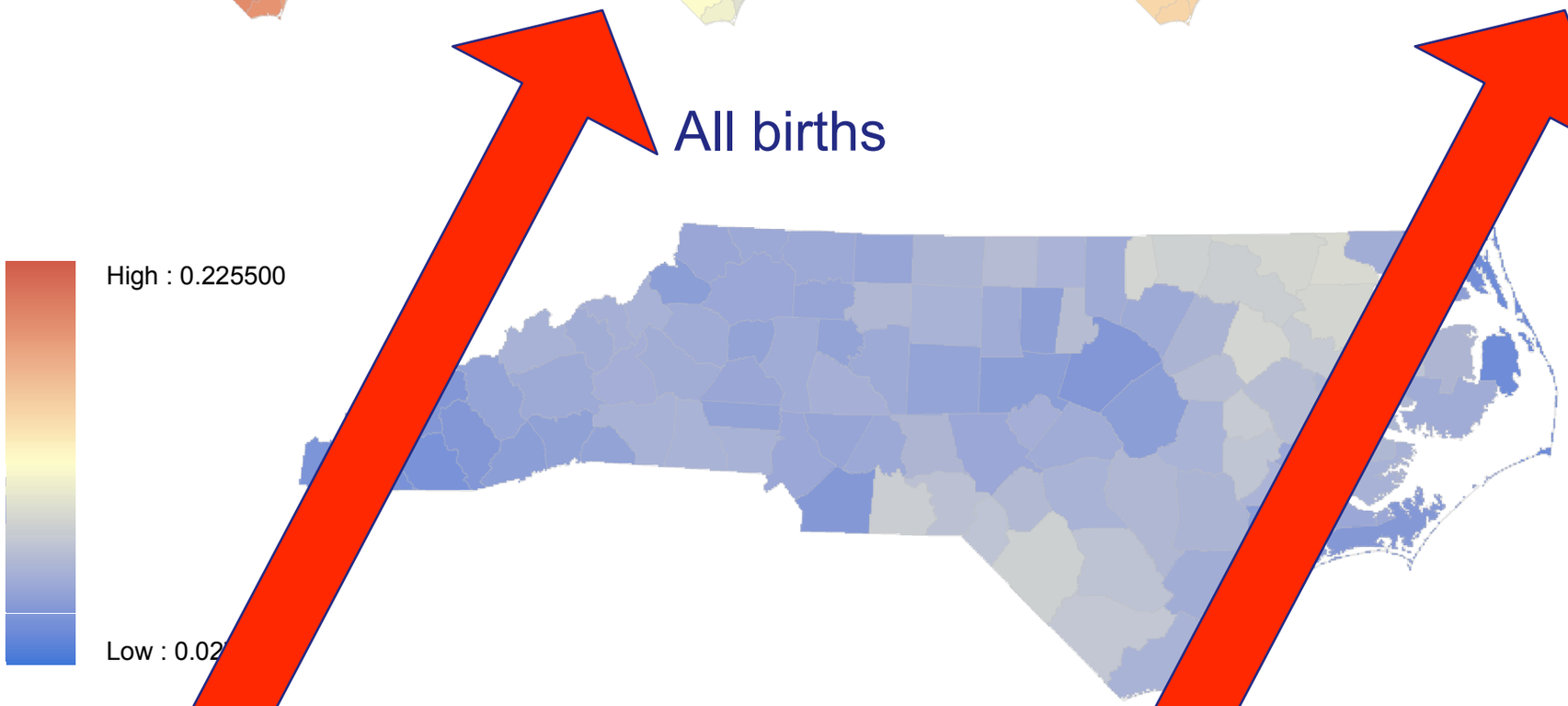
Boy, AA Smoker



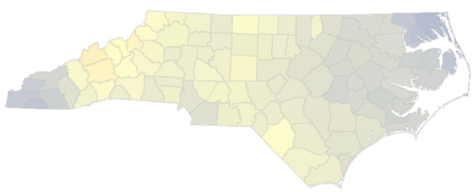
Boy, AA nonSm



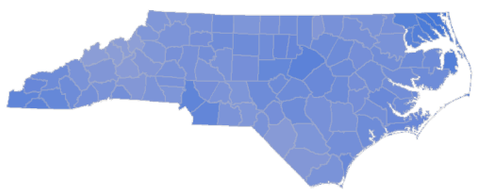
All births



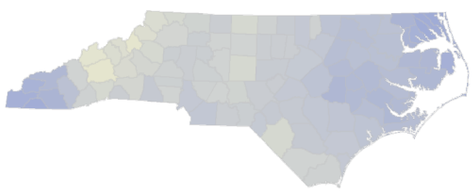
Girl, W nonSm



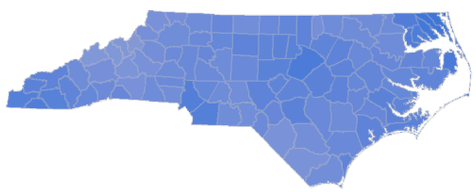
Girl, W nonSm



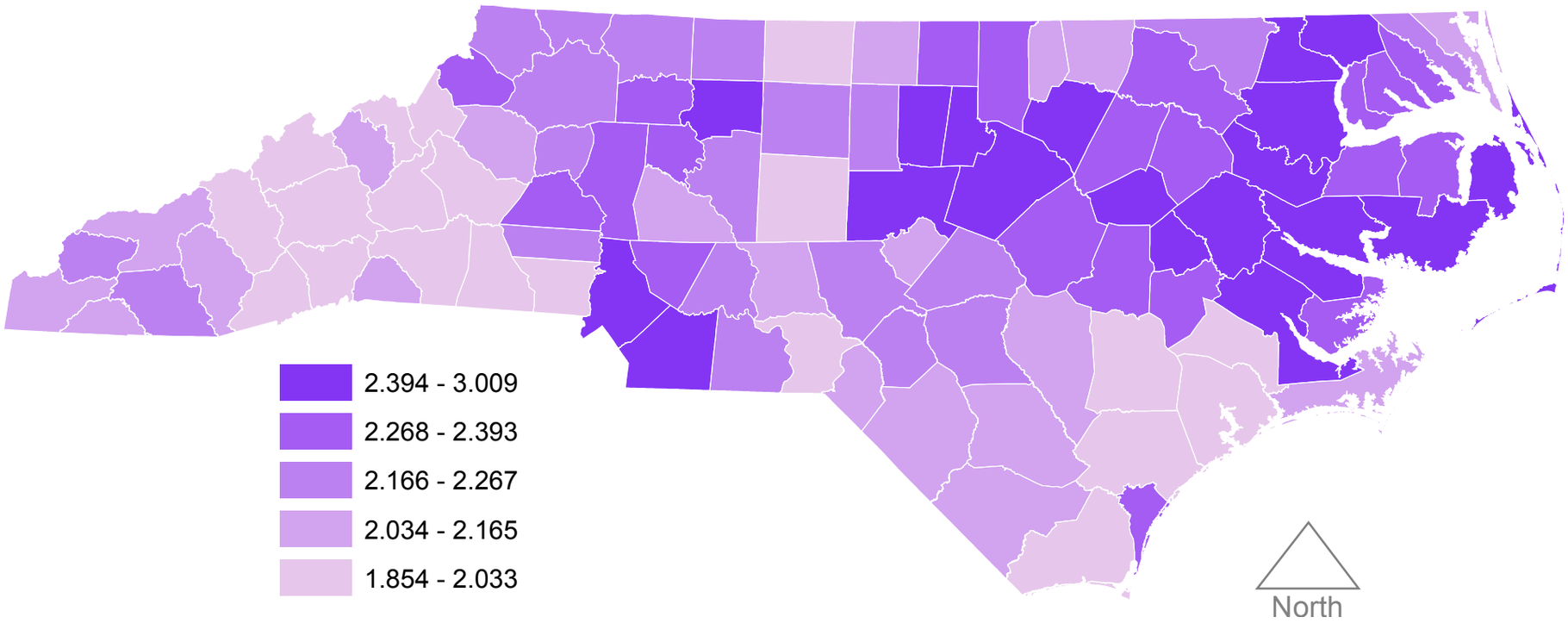
Boy, W nonSm



Boy, W nonSm



Odds Ratio for Race



We can think of this as a (relative) measure of racial disparity.

Impact on Policy?

- Healthy People 2010 (HP2010)
  - "Eliminate Health Disparities"
    - One of two overarching goals

  - HP2010: Health disparities are "differences that occur by gender, race or ethnicity, education or income, disability, geographic location, or sexual orientation."

- Today's example:
  - (1) gender (i.e., subgroups),
  - (2) race (i.e., subgroups), and
  - (3) geographic location (i.e., spatial)
    - ... and combines, in local disparity measures

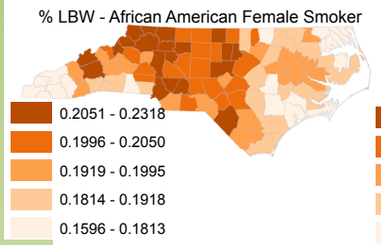
Impact on Policy?

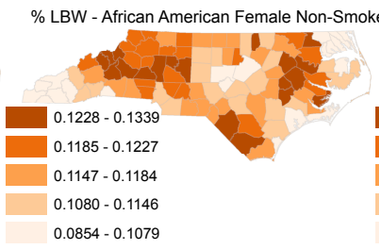
- Also, contextual effects via multilevel model framework
  - E.g., does 'individual' effect of race differ in areas w/ different socioeconomic or demographic features
- Measurement of health disparities
  - "Methodological Issues in Measuring Health Disparities", NCHS (2005)
  - Work of Harper and Lynch
- "Methodological Issues..." emphasize both <u>absolute</u> and <u>relative</u> measures of disparity.
  - OR for race, shown above, is relative measure...
  - ...but flexible methods such as ours can easily accommodate both in some model
  - Also estimates component rates simultaneously

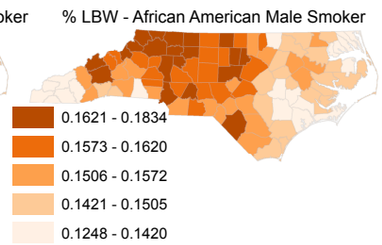
Impact on Policy?

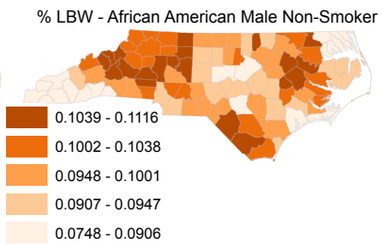
Helpful to see disparity in the context of the component rates...

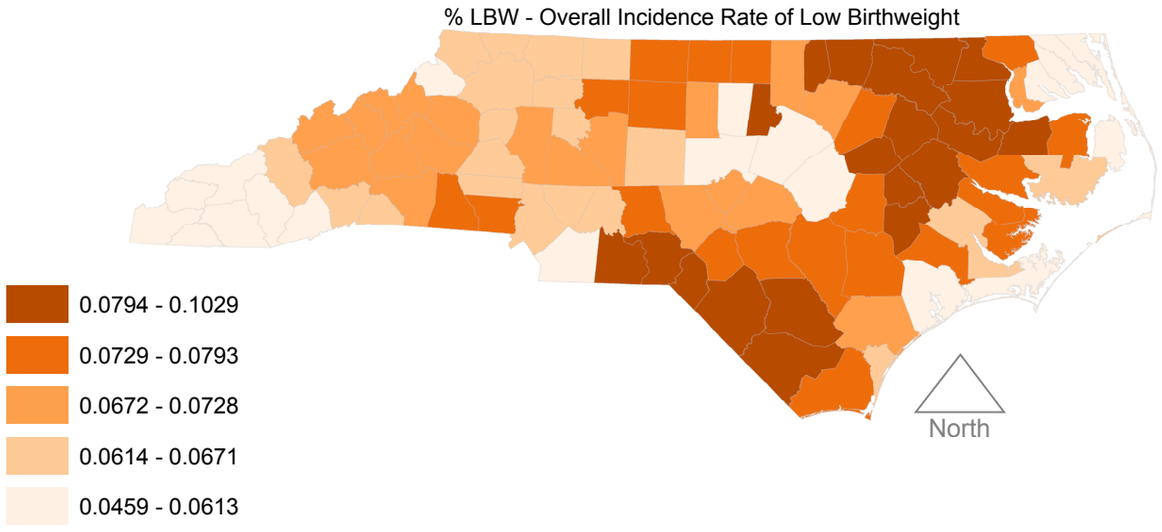
LBW, all and subgroups

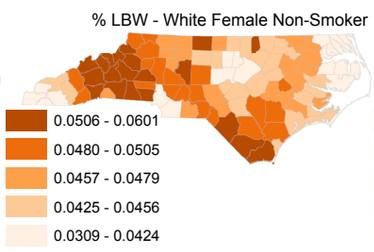


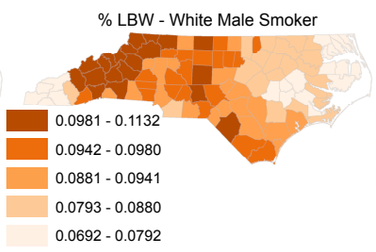


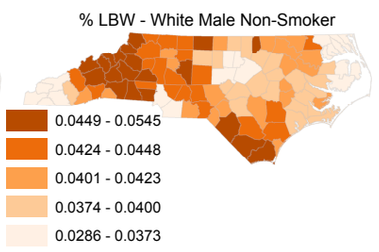












OR race

Impact on Policy?

- How might this inform policies?
  - Different priorities might follow from different relative or absolute disparity… which is higher priority?
  - Different "types" of disparity might might suggest different interventions…
    - How does "two subgroups doing relatively well, but disparity high" compare to "both subgroups doing relatively poorly but disparity low"?
    - Focused intervention in area where one group doing relatively well but other relatively poorly?
- Alternative measures of disparity? Excess deaths?
- Proper measuring of interventions…what if both groups rate's go down, but disparity worsens? Suggests 'disparity only' measures miss something…

Future Directions

- More individual-level data / more categories
  - E.g., maternal education, an ordered categorical variable
  - Logic functions to reduce dimension
- More contextual variables (areal unit covariates)
  - income, demographics, etc., …
  - interact w/ ind. level variables? Variables "with themselves"…

Future Directions

- Spatial scale
  - Now: Counties,
  - Next: ZCTA and beyond...
- Spatial loglinear modeling for point data...

Future Directions

- Spatial scale
  - Now: Counties,
  - Next: ZCTA and beyond…

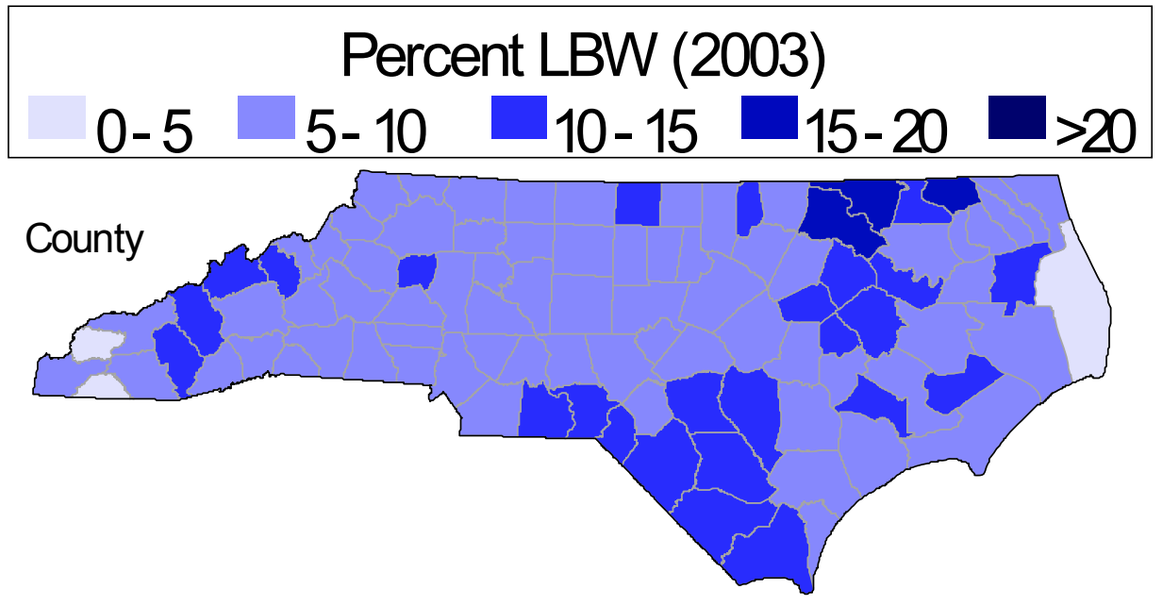
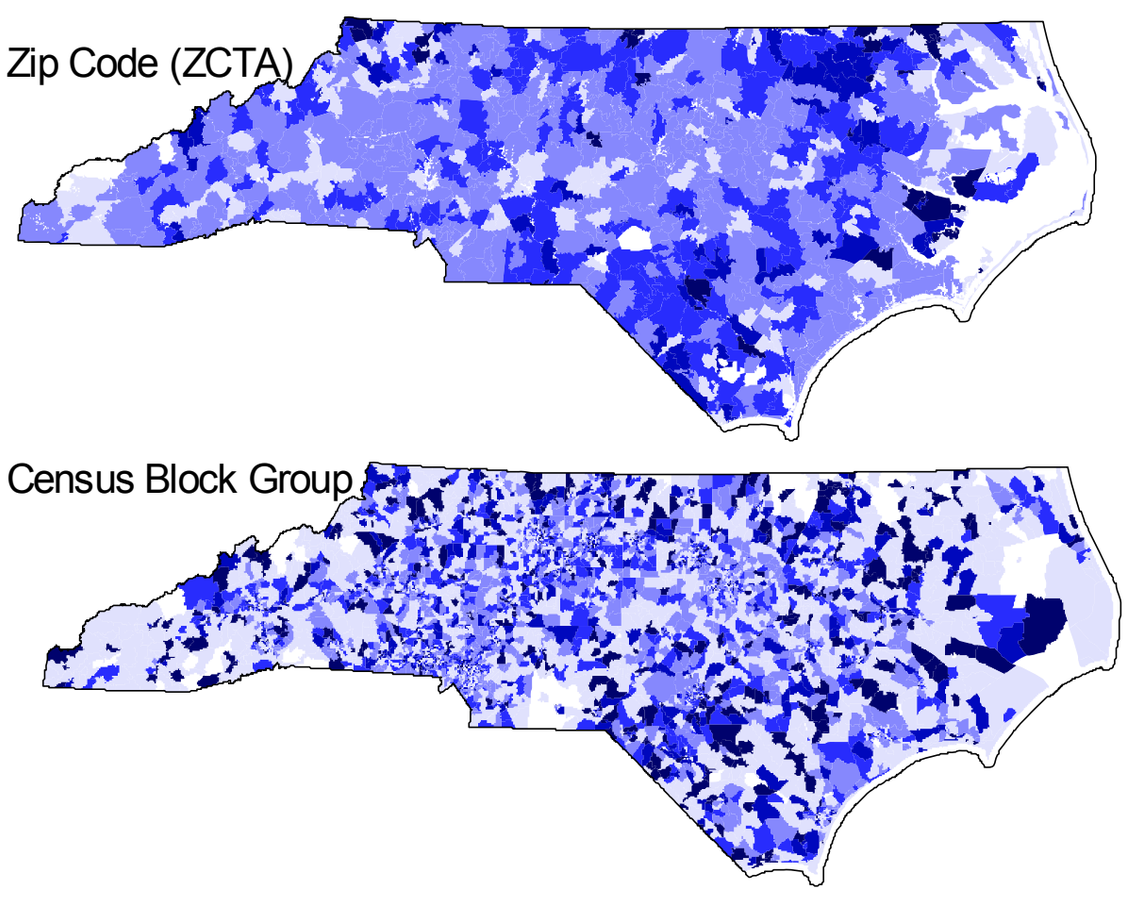


Figure 1. Spatial pattern in percent of low birthweight births in North Carolina.

- Spatial loglinear modeling for point data…



- # Thank you very much!