Community socioeconomic deprivation and obesity trajectories in children using Big Data

Presentation at the Annual Meeting of the Society of Epidemiologic Research, Boston MA, June 18-21 (Spotlight session)

C. Nau¹, T.A. Glass¹, Karen Bandeen-Roche¹, A.Y. Liu¹, J. Pollak¹, A. Hirsch², L. Bailey-Davis², B.S. Schwartz¹,²

¹ Johns Hopkins Global Center on Childhood Obesity, Bloomberg School of Public Health, Baltimore MD
² Center for Health Research, Geisinger Health System, Danville, PA
Motivation

- Overweight in children linked to community-level socioeconomic deprivation
  - Cross-sectional studies only

- Most studies ...
  - ...cannot assess
    - Contextual effects on growth trajectories
  - ... because the studies used
    - Narrow geographies (mostly urban)
    - Little to no longitudinal information
    - Not enough cases per community (e.g. NHANES)
1. Is community socioeconomic deprivation associated with body mass index in children?

2. Are body mass index trajectories modified by community socioeconomic deprivation?

3. Is the association of community socioeconomic deprivation independent of low household income?
Big Data: Electronic Health Records from Geisinger Health System 2000-2010

- N = 161,771
- Age: 3-18 years
- Average of 3.2 obs/child, Range = 1-13
- Living in 1289 communities in 37 Pennsylvania counties
- Average of 127 children/ community Range = 1- 2631
- Mixed definition of place: used minor civil division boundaries for townships and boroughs and census tract in cities
Measure of Community Socioeconomic Deprivation Confirmatory Factor Analysis

Factor analysis

- Less than high school educ.
- Civilian labor force unemployed
- Not in labor force
- Households below poverty
- Households receiving public assistance
- Households with no car

Community socioeconomic deprivation

Histogram of CSD 2000

N: 1289
Min value: -7.8
Max value: 23.7
Mean: 0.00
Std. dev.: 4.3
Methods: Key Measures

- **Time varying:**
  - Body mass index (measured weight [kg] / height [m]^2)
  - Age (years at t)

- **Time invariant:**
  - Sex (1 = female)
  - Race/ethnicity (Hispanic, Black, Other, White [reference])
  - Medical Assistance (1 = enrolled in Medicaid/means tested health insurance for multiple visits)
Statistical Analysis

- **Outcome**
  - BMI instead of BMIz

- **Hierarchical mixed effects models**
  - Level 1: repeated measures of BMI over time
  - Level 2: child level covariates

- **Complex role of age**
  - age, age^2, age^3 (interactions with sex & Medical Assistance)

- **Random effects**
  - Child-level intercept
  - age, age^2

- **Non-stationarity of residual errors**
  - Separate L1 residuals, for 3 age groups

- **SAS Proc Mixed**
## Results: Fixed Effects

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Estimate</td>
<td>Estimate</td>
</tr>
<tr>
<td>Age (centered)</td>
<td>0.8932 ***</td>
<td>0.8393 ***</td>
<td>0.8157 ***</td>
</tr>
<tr>
<td>Age²</td>
<td>0.0198 ***</td>
<td>0.0199 ***</td>
<td>0.0199 ***</td>
</tr>
<tr>
<td>Age³</td>
<td>-0.0052 ***</td>
<td>-0.0051 ***</td>
<td>-0.0047 ***</td>
</tr>
<tr>
<td>CSD 2nd Quartile</td>
<td>0.1319 ***</td>
<td>0.4567 ***</td>
<td>0.4235 ***</td>
</tr>
<tr>
<td>CSD 3rd Quartile</td>
<td>0.1778 ***</td>
<td>0.5855 ***</td>
<td>0.5279 ***</td>
</tr>
<tr>
<td>CSD 4th Quartile</td>
<td>0.2290 ***</td>
<td>0.6968 ***</td>
<td>0.5922 ***</td>
</tr>
<tr>
<td>Age x CSD 2nd Qtle</td>
<td></td>
<td>0.0618 ***</td>
<td></td>
</tr>
<tr>
<td>Age x CSD 3rd Qtle</td>
<td></td>
<td>0.0653 ***</td>
<td></td>
</tr>
<tr>
<td>Age x CSD 4th Qtle</td>
<td></td>
<td>0.0790 ***</td>
<td></td>
</tr>
<tr>
<td>Age² x CSD 2nd Qtle</td>
<td>-0.0004</td>
<td>-0.0010</td>
<td></td>
</tr>
<tr>
<td>Age² x CSD 3rd Qtle</td>
<td>-0.0016 **</td>
<td>-0.0025 ***</td>
<td></td>
</tr>
<tr>
<td>Age³ x CSD 4th Qtle</td>
<td>-0.0002</td>
<td>-0.0015 *</td>
<td></td>
</tr>
<tr>
<td>Age³ x CSD 2nd Qtle</td>
<td>-0.0003 ***</td>
<td>-0.0003 **</td>
<td></td>
</tr>
<tr>
<td>Age³ x CSD 3rd Qtle</td>
<td>-0.0002 **</td>
<td>-0.0002 *</td>
<td></td>
</tr>
<tr>
<td>Medical Assistance</td>
<td></td>
<td></td>
<td>0.5306 ***</td>
</tr>
<tr>
<td>Age x Medical Assistance</td>
<td></td>
<td></td>
<td>0.1133 ***</td>
</tr>
<tr>
<td>Age² x Medical Assistance</td>
<td></td>
<td></td>
<td>0.0027 ***</td>
</tr>
<tr>
<td>Age³ x Medical Assistance</td>
<td></td>
<td></td>
<td>-0.0009 ***</td>
</tr>
<tr>
<td>-2 LogLikelihood</td>
<td>2230103</td>
<td>2229762</td>
<td>2228833</td>
</tr>
</tbody>
</table>

Notes: *** p-value < 0.001, ** p-value <0.5, * p-value <.1. All models control for sex, age*sex interactions and race.
RESULTS
Community Socioeconomic Deprivation

Cubic Regression Spline by C-SED Quartile

Partial residuals

Centered Age in Years

4 quantiles of econdep

-4 -3 -2 -1 0 1 2 3 4 5 6 7 8

-3 -2 -1 0 1 2 3
Big Data from Electronic Health Records

- **Disadvantages**
  - Computational challenges to estimate complex models on very large data sets
  - Limited information on individual/family socioeconomic status
  - Limited diversity by race/ethnicity

- **Advantages**
  - Large sample size
  - Large spatial and temporal coverage
  - Height and weight measured consistently in a clinical setting
  - Routine collection of repeated measures
  - Detailed medical history
  - Possibility to control for multiple potential confounders
Conclusion

1. Residence in socioeconomically deprived communities is associated with obesogenic growth trajectories
   - Age associations vary by CSD
   - Evidence of exposure-effect relation

2. Community socioeconomic deprivation association is independent of confounding influence of race/ethnicity, sex, and a surrogate for low family income

3. Big Data from electronic health records facilitates study of contextual effects on growth trajectories
Project 1: Dynamics of childhood obesity in Pennsylvania from community to epigenetics

Johns Hopkins Project Team:
- Thomas A. Glass (Epi), Project leader
- Brian S. Schwartz (EHS), Project co-leader
- Karen Bandeen-Roche (Bio), Biostatistician
- Joseph Bressler (EHS), epigenetics
- Ann Liu (EHS), Research associate
- Tak Igusa (Engineering), Systems scientist
- Claudia Nau (IH), Post-doctoral fellow, trainee
- Mehdi Jalalpour (Eng), Pre-doctoral fellow, trainee
- Jonathan Pollack (EHS), Data analyst

Geisinger Health System Team:
- Annemarie Hirsch, subcontract PI
- Lisa Bailey-Davis, co-investigator
- Dione Mercer, Project coordinator
- Sy Landau, Research assistant
- Joseph DeWalle, GIS analyst

Statement of funding support:
This research was funded in part through a cooperative agreement from the National Institutes of Health (U54 HD070725). Contributing partners include the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) and the Office of Behavioral and Social Sciences Research (OBSSR).
END OF PRESENTATION