ONEdata
Obesity &
Neighborhood Environment
Database

Misa Graff, Ph.D.
On behalf of the ONEdata Group
University of North Carolina at Chapel Hill
Purpose of ONEdata: Characterize the Obesogenic Neighborhood
Obesity & Environment

- Research findings
- Database development process
- Overview of environment data
Obesity and Environment Research Findings
Background

- Adolescence and young adulthood are major periods of biological, social and behavioral development, with potential importance for future health behaviors.

- There is a paucity of research on the multiple dimensions of influence operating on behaviors during adolescence and young adulthood:
  
  - Household, school, community and wider environment?

- This is what prompted us to generate the ONEdata Database.
An Ecological Model of Diet, Physical Activity, and Obesity

Influences
- Biological & Demographic
- Psychological
- Social/Cultural
- Organizational
- Physical Environment
- Policies/Incentives

Behaviors
- Eating Diet, nutrition
- Sedentary Behaviors
- Physical Activity

Energy Balance

Health Outcomes
- Body Weight, Fat, & Distribution
- Risk Factors, CVD, Diabetes, Cancers

Developed for the NHLBI Workshop: Predictors of Obesity, Weight Gain, Diet, and Physical Activity; August 2004, Bethesda MD
The National Longitudinal Study of Adolescent Health (Add Health)

♦ Wave I (1995) - 20,745 respondents
♦ Wave II (1996) – 14,738 wave I respondents (in school)
♦ Wave III (2001) – 15,197 original wave I respondents
♦ Wave IV (2007) - ~17,000 original wave I respondents
ONEdata includes environment data at two waves

- Wave I (1995) - 20,745 respondents
- Wave II (1996) – 14,738 wave I respondents (in school)
- Wave III (2001) – 15,197 original wave I respondents
- Wave IV (2007) - ~17,000 original wave I respondents
BMI Distribution of Add Health respondents at adolescence and young adulthood

Proportion of respondents with ≥5 bouts MVPA/wk drops from adolescence & adulthood

Wave I, 1995 (adolescence)  
Wave III, 2001 (early adulthood)

Proportion of respondents with ≤14 hours of weekly “screen time” remains relatively high.

Research Question

♦ What about modifiable environmental factors?
There Are Important Associations Between Modifiable Environmental Factors and Activity Patterns of US Adolescents

Adjusted for sex, age, SES, urban residence, in-school status, pregnancy, region and month of interview

Gordon-Larsen, et al. 2001; Pediatrics
Research Question

♦ At a national level: Are physical activity resources and facilities equitably distributed by SES and race/ethnicity?

♦ How does distribution of resources impact health outcomes?
Add Health Neighborhoods
42,857 block groups w/in 5 mi of respondent

Individual Buffer
Respondents
Sampled Block Group
School
Community Study Area

Obesity & The Environment
The University of North Carolina at Chapel Hill
Add Health Neighborhoods
42,857 block groups w/in 5 mi of respondent

Respondents
Individual Buffer
Sampled Block Group
Block Group Boundary

Community Study Area
School
Resources

Obesity & The Environment
The University of North Carolina at Chapel Hill
Aggregation of Physical Activity Facilities [N=67,080]
From Digitized Business Records

- **Schools**: Elementary, secondary, college, university
- **Public**: Public swimming pools, tennis courts, parks
- **Youth organization**: Boy/girl scouts, youth centers,
- **Parks**: Park and recreation services
- **YMCA**: YMCA, YWCA
- **Public Fee**: Physical fitness facilities, bicycle rental
- **Instructional**: Activity-related classes or instructors
- **Outdoor**: Sporting and recreational camps, swimming pools
- **Member**: Athletic club and gymnasium, tennis club
- **All**: All facilities across all categories

*Obesity & The Environment*
*The University of North Carolina at Chapel Hill*
Analysis Methods

Logistic regression analyses tested:

- The relationship of PA-related facilities with block group socioeconomic status (SES)

(All analyses control for population density)
Odds of Having at Least One PA Facility are Higher Among Neighborhoods With More Educated† Populations‡

† Increased odds given each 100% increase in population with college education+
‡ Adjusted by population density and % minority population

Odds of Having at Least One PA Facility are Higher Among Neighborhoods With More Educated† Populations‡

<table>
<thead>
<tr>
<th>Category</th>
<th>Adjusted Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Facilities</td>
<td>2.18</td>
</tr>
<tr>
<td>Public Schools</td>
<td>2.15</td>
</tr>
<tr>
<td>Youth Organizations</td>
<td>2.84</td>
</tr>
<tr>
<td>Parks</td>
<td>2.97</td>
</tr>
<tr>
<td>YMCA</td>
<td>2.57</td>
</tr>
<tr>
<td>Instruction</td>
<td>4.2</td>
</tr>
<tr>
<td>Outdoor Member</td>
<td>4.22</td>
</tr>
<tr>
<td>Member Fee</td>
<td>1.41</td>
</tr>
</tbody>
</table>

† Increased odds given each 100% increase in population with college education+
‡ Adjusted by population density and % minority population

Odds of Having at Least One PA Facility are Higher Among Neighborhoods With More Educated† Populations‡

† Increased odds given each 100% increase in population with college education+
‡ Adjusted by population density and % minority population

Lowest Education and Highest Minority Population Were Least Likely to Have at Least One PA Facility *

Odds Ratios (95% CI) include minority*education interaction, adjusting for population density

Logistic regression analyses tested:

- Subsequent association of PA facilities with likelihood of:
  - being overweight (BMI $\geq 95^{th}$)
  - achieving 5+ bouts/week moderate-vigorous PA

(Control for population density)
## Odds of Overweight Decreases and MVPA Increases with Greater Number of Facilities

<table>
<thead>
<tr>
<th>Recreation Facilities (#)</th>
<th>Adjusted OR (95% CI)</th>
<th>Adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overweight</td>
<td>MVPA</td>
</tr>
<tr>
<td>1</td>
<td>0.95 (0.90-0.99)</td>
<td>1.03 (1.0-1.06)</td>
</tr>
<tr>
<td>2</td>
<td>0.90 (0.82-0.98)</td>
<td>1.07 (1.02-1.21)</td>
</tr>
<tr>
<td>3</td>
<td>0.85 (0.74-0.97)</td>
<td>1.10 (1.03-1.19)</td>
</tr>
<tr>
<td>4</td>
<td>0.80 (0.67-0.96)</td>
<td>1.14 (1.03-1.26)</td>
</tr>
<tr>
<td>5</td>
<td>0.76 (0.60-0.95)</td>
<td>1.18 (1.04-1.33)</td>
</tr>
<tr>
<td>6</td>
<td>0.72 (0.55-0.95)</td>
<td>1.22 (1.05-1.41)</td>
</tr>
<tr>
<td>7</td>
<td>0.68 (0.49-0.94)</td>
<td>1.26 (1.06-1.26)</td>
</tr>
</tbody>
</table>

Odds Ratios (95% CI) adjusting for population density

Summary of Findings

♦ At a national level all major categories of physical activity-related resources are inequitably distributed

♦ Low SES, minority neighborhoods at strong disadvantage

♦ Further, this inequitable distribution is significantly associated with subsequent disparities in health outcomes
What does Add Health data offer?

♦ Provides unprecedented opportunity to explore determinants operating at multiple levels of influence
♦ Large samples of ethnically diverse adolescents followed over time, including special subsamples, such as siblings
♦ Unique environmental database offers exciting potential for research
Obesity and Environment Database Development Process
The Obesogenic Environment
Obesogenic environment
Obesogenic environment

Dan Burden, Walkable Communities Inc.
Obesogenic environment

Dan Burden, Walkable Communities Inc.
Obesogenic environment
Suburban development, many cul-de-sacs

Well-connected urban development with mixed land uses
Environment can be broadly defined when looking at obesity, activity, and diet

- **Built Environment**
  - Urban design, land use, transportation system
  - Sprawl, walkability

- **Economic Context**
  - Prices: Housing, cost of living

- **Sociodemographic context**
  - Community race/ethnicity, income, wealth

- **Social context**
  - Crime, traffic, aesthetics, degradation

- **Sociopolitical context**
  - Zoning, governance, legal realm
Definition of Built Environment

♦ “All buildings, spaces and products that are created, or modified, by people.
  • It includes homes, schools, workplaces, parks/recreation areas, greenways, business areas and transportation systems.”

♦ Urban design, land use, transportation systems

♦ “Consists of the neighborhoods, roads, buildings, food sources, and recreational facilities in which people live, work, are educated, eat, and play”

(RFA-ES-04-003, Sallis & Glanz 2006)
Place Locations onto Map
Geographic Information System (GIS)

- GIS is essentially a computerized map
  - allows plotting of resource layers onto a coordinate system
  - Can then spatially analyze the density and proximity of resources, environment factors, and population characteristics
- Using GIS, locations can be geocoded, or assigned a geographic reference, such as latitude and longitude.
For a GIS approach

- Have addresses or GPS for each respondent
- Geocode these addresses
- Build environment database
- Join environment database to individual attribute data
Building the database

Used 2 time points: Waves 1 and 3
Integrate built environment measures into a database that can then be linked to individual-level attribute data

- Linkage to contextual databases through collecting detailed location data by street address GIS and GPS
- Linkage to broad set of national data on: Economics, Policy, Zoning, Government Regulation at small levels of geographic scale
- Enable ability to get at multiple levels of influence
  - individual family/home, school/peers, community, industry/government, culture/society
State of GIS-derived research

- National samples using large geographic scale units (e.g., state-level effects)
- Highly detailed measures, sometimes collected via audits, in one small geographic area
ArcView & ArcGIS software are not designed for population-based studies

- Problem: ArcView & ArcGIS GUI interfaces designed for user-driven “one-off” analyses and operations
  - Requires building software to run on top of ArcView & ArcGIS to drive the data processing flow

- Problem: ArcView & ArcInfo have database size limitations exceeded by Add Health national sample
  - Requires coding workarounds or developing custom software alternatives (e.g. Net-Engine, Python, Avenue, C++)
Obesity and Environment
Overview of Environment Data
Respondent Locations

- Respondent residential locations geocoded
  - Street-segment matches from address geocoding given precedence (83% match rate)
  - GPS (15% match rate) and ZIP+4/ZIP+2/ZIP matches (2% match rate) used to “fill in”
Unit of Analysis, Geographic Scale (varies across source data)

- Administrative boundaries
- Buffers
  - Circular
  - Polygon-based road network buffers

Oliver LN, Schuurman N, Hall AW. Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. Int J Health Geogr. 2007 Sep 20;6(1):41
Building the Obesity and Environment Database Add Health

ורודנים קהילות, 80 מחוזות, 33 מדינות, >42,000 קבוצות בלוק

 punching

 Challenges took 7 years of research effort by our UNC group to overcome

- national scale: requires significant GIS programming skills
- scale and coordinate conversion issues: aligning databases spatially and temporally
- security and confidentiality
Building the Obesity and Environment Database
Add Health

- Need “industrial-strength” GIS tools to handle data volume
- Must be customizable with scripting/programming languages

- Environmental Systems Research Institute (ESRI)
  - ArcInfo 8.x-9.x (customized with AML)
  - ArcView 3.3 (customized with Avenue)
  - ArcGIS 8.x-9.x (customized Python & Visual Basic)
  - NetEngine (customized with C)
It takes a village of trained professionals
Data Steps

Geocoding

- QA/QC respondent locations

Build 8.03 km buffers

- Built buffer for each respondent

Data Evaluation Project

- Determine which external datasets had appropriate spatial/temporal data and accuracy (Dynamap) and create subset datasets corresponding to each community study area

Spatial Join

- Merge subset datasets corresponding to each community study area

QA/QC

- Evaluation
### Join database to individual-level data

<table>
<thead>
<tr>
<th>ID</th>
<th>BMI</th>
<th>BG: % in poverty</th>
<th>Number of Parks</th>
<th>Number of McDonalds</th>
</tr>
</thead>
<tbody>
<tr>
<td>10090</td>
<td>31.5</td>
<td>43.2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>10091</td>
<td>25.5</td>
<td>23.4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>10000</td>
<td>28.9</td>
<td>19.1</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>10123</td>
<td>38.2</td>
<td>2.5</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Some validation work

- Physical activity facilities database
Characteristics of the validation locations in two non-Add Health communities

<table>
<thead>
<tr>
<th></th>
<th>Non-Urban mean^† (SD) (n=40)</th>
<th>Urban mean^† (SD) (n=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block group area (mile(^2))</td>
<td>3.7 (8.6)</td>
<td>0.03 (0.02)</td>
</tr>
<tr>
<td>Block group population</td>
<td>1,753 (948)</td>
<td>1,213 (782)</td>
</tr>
<tr>
<td>Population density (persons/ mile(^2))</td>
<td>1,634 (1,152)</td>
<td>58,581 (35,285)</td>
</tr>
<tr>
<td>White non-Hispanic population</td>
<td>960 (673)</td>
<td>815 (768)</td>
</tr>
<tr>
<td>Median household income</td>
<td>$40,157 (21,794)</td>
<td>$33,925 (22,059)</td>
</tr>
</tbody>
</table>

^†unweighted average among block groups, using 2000 Census data
Example of potential influence of GIS error on counts of facilities
Agreement* of *administratively defined* neighborhood and nearest street of GPS and geocoded physical activity facility locations

<table>
<thead>
<tr>
<th>Community</th>
<th># Facilities</th>
<th>5-Digit ZIP Code</th>
<th>Census Tract</th>
<th>Block Group</th>
<th>Street</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Urban</td>
<td>63</td>
<td>59</td>
<td>59</td>
<td>58</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>41</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Non-Urban    | 94%          | 94%              | 92%          | 71%         |        |
| Urban        | 100%         | 100%             | 98%          | 71%         |        |

*Agreement calculated among facilities in both the GIS and field census.

Challenges

Dataset Integration

- Scale and coordinate conversion issues related to our databases and aligning these databases spatially, temporally, and communally
Challenges

Solution

- Painstaking effort of examining record by record and location by location to confirm…
  - all components of the database were spatially aligned
  - all components were accurate, complete, and well linked
- In some cases, this required manual, visual comparison of data against Digital Orthophoto Quads (DOQ)
- A necessary undertaking -- major problems found in multiple source datasets
- Example: Geographic misalignment in one area within underlying street data used for respondent & other locations in one area
  - required spatial adjustment in GIS database.
Challenges

National data development effort \( \rightarrow \) “exception” cases will be encountered!

Example: “Minute triangles” for Census units in TIGER/Line data

- Arbitrary polygonal representation of Census tracts & blocks for crews-of-vessel populations
- Small number nationwide, but neighborhoods for our sample include them
- Skewed neighborhood population density calculations
Challenges

Security and Confidentiality

- need for de-identifying data to ensure confidentiality of respondents
- all file linkage and identifiers maintained by York University in Canada
- limited ability to complete true exploratory spatial analysis

Solutions:

- designate specific personnel and protected hardware to work with location data
  - no linkage for them to respondent data
- project investigators receive only derived data, no location data
It takes a village of trained professionals.
Acknowledgements

♦ Penny Gordon-Larsen (UNC, Nutrition)
♦ Barry Popkin (UNC, Nutrition, Economics)
♦ Linda Adair (UNC, Nutrition)
♦ Kari North (UNC, Epidemiology)
♦ David Guilkey (UNC, Economics)
♦ Yan Song (UNC, City & Regional Planning)
♦ Postdocs: Janne Boone-Heinonen, Misa Graff
♦ Doctoral students: Ningqi Hou, Natalie The, Melissa Nelson, Whitney Robinson Melissa Scharoun-Lee

Spatial Analysis Unit (CPC): Phil Page, Jay Stewart, Marc Peterson, Evan Hammer and others

NIH funding: NICHD, NIDDK, NIEHS
For More Information

♦ http://www.cpc.unc.edu/projects/onedata