The Built Environment, Food, and Physical Activity: Measurement Issues
Ann Forsyth, PhD
Department of City and Regional Planning
Cornell University
January 19, 2009

1. Built Environment and Behavior—Measuring Relevant issues

2. GIS Measures
   - Basic approach, data, variables, and geographies

3. Typical Dilemmas
   - Definitions, data quality, buffering changes

   - My studies 2003-2006 TCWS (RWJF, PI) and cluster of NIH-funded studies from 2006 (IDEA, EAT-III, ECHO, Co-PI/consultant)

1. Built Environment
Start of Recent Interest
- Popular interest in obesity
- Weight = food minus activity
- PA initial focus—early (2002) studies at county level found exercise did not vary by environment but weight did
- Explanations:
  - Research design
  - Food intake
  - Social and cultural factors
  - Non-leisure physical activity
- Travel research had found commute walking varies with environment
- Education not working
- Seemed possible to increase overall activity through utilitarian physical activity (PA)

1. Heart Disease 28.5% 12.4% (#1)
2. Cancer 22.8% 2.2% (#10, lung)
3. Stroke 6.7% 9.2% (#2)
4. Emphysema, chronic bronchitis 5.1% 4.5% (#5)
5. Accidents 6.7% 2.3% (#9, road)*
6. Diabetes 3.7%
7. Flu & pneumonia 2.7% 6.9% (#3)
8. Alzheimer/senility 2.4%
9. Kidney disease 1.7%
10. Septicemia/systemic infection 1.4%

Other top 10 in the world: HIV/AIDS 5.3% (#4), perinatal (#6), diarrhea (#7), tuberculosis (#8) All data for 2002

1. Built Environment
How Active are People Now?
- We have only recently had good large scale studies (e.g. until recently health studies asked about exercise, and travel studies about travel)
- People are active in different parts of their lives
  - Work
  - Exercise and leisure
  - Care/chores/errands
  - Commuting to work
- Still difficult to assess—all collection mechanisms have weaknesses (surveys, diaries, gadgets)

1. Built Environment
Theories about What Matters for Walking (and Biking)
- Transportation
- Urban/landscape design (physical planning, architecture, landscape architecture)
- Physical activity research
1. Built Environment

Interesting New Analysis Methods

**Grid Looped Street lights**

But data are not always available—good to have photos but....

1. Built Environment

The Food Side of the Energy Balance Equation

- Food has similar variety of theories on how environment affects eating
- Is proximity important?
  - Proximity to what (home, work, route), healthy or unhealthy?
  - Food deserts vs. ubiquitous food (name a retail place in Collegetown that does not sell any food)
  - Relative importance of economic, social, and psychological issues in consumption
  - Whether food access affects walking
- Very little known until recently about where people actually shop (vs. what they eat)
- Doesn’t build on rich theoretical/empirical base from spatial disciplines compare food + planning/geography groups with whole fields of transportation, planning, design

What GIS can help measure

* = Mediating factors

* = Within the domain of land use and transportation planners

**Built Environment, Food, and Physical Activity:**

Measurement Issues

1. Built Environment and Behavior—Measuring Relevant issues

2. GIS Measures
- Basic approach, data, variables, and geographies

3. Typical Dilemmas
- Definitions, data quality, buffering changes

4. Analysis
2. Measures

GIS

- A Geographic Information System (GIS) is a system of software and data that allows mapping and analysis
- ESRI is the main vendor in the US (ArcGIS suite including ArcMap, ArcCatalog, Spatial Analyst, ArcInfo, Network Analyst)
- Menu and wizard-based rather than scripted—problems of replication
- Rapid increase in available software and data over the past 10 years—though little collected specifically for research
- Distinctive cultures—remote sensing, planning/urban studies, environmental, social analysis

2. Measures

How to Think About Measures

Measurement of food/activity variables associated with the built environment occurs on several levels:
- Behavior of interest e.g. healthy food intake
- Dimension of interest e.g. restaurant meals (quality)
- Measurement of variable (formula) associated with that dimension e.g. all restaurants, all limited service restaurants, major chains (different data sources limit kinds of definitions)
- Geography of that measurement e.g. 400 meter street network distance/buffer, distance to the nearest something, home to school route
- Data: commercial lists, licensing, land use/parcels, fieldwork

There can be easily 50-100 variable/geography combinations for one dimension or topic e.g. five different restaurant measures at up to 20 different geographies

2. Measures

Data Sources for Food and PA

- Fieldwork (RWJF has sponsored a number of tools)
- Land Use and Parcel Data (including interpretation of aerial photographs)
- Food Licensing Data
- Commercial Business Data
- Business Taxation data
- Online Street-level Photographs—not used yet (example from Google Street View [launched May 2007] in Valencia [added October 2008])

All can be used in GIS
2. Measures

Data Sources for Food and PA

- Fieldwork
- Land Use and Parcel Data (including interpretation of aerial photographs)
- Food Licensing Data
- Commercial Business Data
- Business Taxation data
- Online Street-level Photographs

- All can be used in GIS—all have weaknesses for locating food sources and measuring quality

2. Measures

Example of TREC-IDEA Study

- Identifying Determinants of Eating and Activity in Adolescents (IDEA), Leslie Lytle PI
- Goal: Examine the etiology of adolescent obesity and risk of future cancer using a social-ecological approach and considering possible risk and protective factors at the individual, family, home, school and neighborhood levels
- Study Design
  - Enroll an existing cohort of youth ages 14-18 plus at least one biological parent (n = 323 at baseline)
  - Collect data on the cohort at three time points over a period of 24 months (baseline; 12 months; 24 months)
  - Primary outcome: Predictors of change in BMI trajectories/weight status categories over time

2. Measures

TREC-IDEA Study

- Measurements planned:
  - Child/Adolescent level: biological markers, body comp, diet, activity, psychosocial
  - Parent: body comp, diet, activity, psychosocial
  - Home: Home food inventory, physical activity and media inventory, meal screener
  - School: Availability of foods/activity; food and activity policies
  - Neighborhood: GIS- activity and eating opportunities around homes and schools
2. Measures

**BE Variables In TREC/IDEA (45 + census)**

- Food related variables (distance to + density of)
  - Restaurants (all, fast food, fast food sub-types, non-fast food [8 measures])
  - Food stores (all, groceries, large supermarkets, convenience stores, nearest type, farmers market, alcohol)
- Physical activity related variables include
  - Density measures
  - Street pattern/connectivity
  - Pedestrian infrastructure
  - Mixed use/destinations including distances to recreation facilities (parks, rec. centers, trails etc)
- Traffic
- Census variables
- Also comparing survey (perceived/behavior) vs. measured environmental features

**TREC Geographies**

- Home straight line/network buffers at 400, 800, 1600, and 3,000 meters (density)
- Home straight line/network distances to nearest feature
- School straight line/network buffers
- School straight line/network distances to nearest feature
- Home to school network buffer (100m either side)
- Each person has up to 21 separate geographies for each variable
- Manage the Modifiable Areal Unit Problem (zoning + scale)

3. Dilemmas: Definitions

**How Food Stores and Restaurants are Defined in the Literature to Date**

<table>
<thead>
<tr>
<th>Data Source (down)</th>
<th>Researcher defined categories</th>
<th>Land use type</th>
<th>Chain names</th>
<th>Industrial code (e.g. NAICS)</th>
<th>Proprietary/ local codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fieldwork</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>May not be specific</td>
<td></td>
</tr>
<tr>
<td>Land use and parcel data</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Food licensing data</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Commercial business data</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

Potential to use online street level photographs e.g. Google StreetView, Tax data

**Example Land Use Data Categories Metro-Wide—Local Data Has More Detail**

- 100 = Agricultural
- 111 = Farmstead
- 112 = Seasonal/Vacation
- 113 = Single Family Detached
- 114 = Single Family Attached
- 115 = Multifamily
- 116 = Manufactured Housing
- 120 = Retail and Other Commercial
- 130 = Office
- 141 = Mixed Use Residential
- 142 = Mixed Use Industrial
- 143 = Mixed Use Commercial and Other

- 151 = Industrial and Utility
- 153 = Extractive
- 160 = Institutional
- 170 = Park, Recreational, or Preserve
- 173 = Golf Course
- 201 = Major Highway
- 202 = Railway
- 203 = Airport
- 210 = Undeveloped
- 220 = Water

Generalized Land Use Categories: Twin Cities Metropolitan Area, 2005

Source: Metropolitan Council 2005
3. Dilemmas: Definitions

Example Chain Names

- Chain names
  - Problem of choosing a list
  - Wikipedia has a long list of over 150 fast food names, also extensive list of supermarkets; most scholars have much shorter lists

A. NAICS codes

- Finding fast food in commercial data sets?
  - Search using North American Industrial Classification System/Standard Industrial Classification System codes (NAICS/SIC codes)
  - 722110 Full-Service Restaurants
  - 722211 Limited-Service Restaurants
  - 722212 Cafeterias
  - 722213 Snack and Nonalcoholic Beverage Bars
  - Etc.
  - Problem of mis-classification (e.g. in D&B Culver's and Famous Dave's are considered full-service restaurants because employees bring food to your table, even though you order at the counter)

http://www.census.gov/epcd/naics02/naicod02.htm

(Mis) Classification

<table>
<thead>
<tr>
<th>Comparison of NAICS Code Distribution for 50 Zip Codes</th>
<th>Business Analyst</th>
<th>Dun and Brad.</th>
<th>Diff %</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAICS Code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>445 Food and Beverage Stores</td>
<td>452</td>
<td>394</td>
<td>-13%</td>
</tr>
<tr>
<td>4451 Grocery Stores</td>
<td>242</td>
<td>218</td>
<td>-10%</td>
</tr>
<tr>
<td>44511 Supermarkets and Other Grocery (except Convenience) Stores</td>
<td>117</td>
<td>70</td>
<td>-40%</td>
</tr>
<tr>
<td>4452 Specialty Food Stores</td>
<td>119</td>
<td>82</td>
<td>-31%</td>
</tr>
<tr>
<td>44521 Meat Markets</td>
<td>17</td>
<td>16</td>
<td>6%</td>
</tr>
<tr>
<td>44522 Fish and Seafood Markets</td>
<td>2</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>44523 Fruit and Vegetable Markets</td>
<td>12</td>
<td>2</td>
<td>-83%</td>
</tr>
<tr>
<td>44529 Other Specialty Food Stores</td>
<td>88</td>
<td>59</td>
<td>-33%</td>
</tr>
<tr>
<td>445291 Baked Goods Stores</td>
<td>0</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>445292 Confectionery and Nut Stores</td>
<td>12</td>
<td>13</td>
<td>8%</td>
</tr>
<tr>
<td>445299 All Other Spec. Food Stores</td>
<td>76</td>
<td>28</td>
<td>-66%</td>
</tr>
</tbody>
</table>

3. Dilemmas: Definitions

(Mis)Classification Part 2

<table>
<thead>
<tr>
<th>Comparison of NAICS Code Distribution for 50 Zip Codes</th>
<th>Business Analyst</th>
<th>Dun and Brad.</th>
<th>Diff %</th>
</tr>
</thead>
<tbody>
<tr>
<td>722 Food Services and Drinking Places</td>
<td>1004</td>
<td>950</td>
<td>-5%</td>
</tr>
<tr>
<td>7221 Full-Service Restaurants</td>
<td>34</td>
<td>432</td>
<td>117%</td>
</tr>
<tr>
<td>7222 Limited-Service Eating Places</td>
<td>829</td>
<td>371</td>
<td>-55%</td>
</tr>
<tr>
<td>722212 Cafeterias</td>
<td>5</td>
<td>8</td>
<td>60%</td>
</tr>
<tr>
<td>722213 Snack and Nonalcoholic Beverage Bars</td>
<td>75</td>
<td>1</td>
<td>-99%</td>
</tr>
<tr>
<td>7224 Drinking Places (Alcoholic Beverages)</td>
<td>91</td>
<td>103</td>
<td>13%</td>
</tr>
</tbody>
</table>

Comparing commercial numbers, correlation is high \( r = 0.96 \) but lower comparing commercial data to licensing numbers \( r = 0.7 \)

3. Dilemmas: Data Quality

Generic Data Problems

- Consistency across jurisdictions
- Purpose—original collection
- Geographies of data collection
- Resolution/scale
- Accuracy
- Completeness
- Reliability information
- Time of collection
- Errors
- Cost—cleaning, coding

<table>
<thead>
<tr>
<th>Data Quality Issue</th>
<th>Fieldwork</th>
<th>Land Use + Parcel Data</th>
<th>Business Licensing</th>
<th>Commercial Listing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistency across jurisdictions</td>
<td>Flexible</td>
<td>Limited</td>
<td>Limited</td>
<td>A strength of many of these data</td>
</tr>
<tr>
<td>Purpose, original collection</td>
<td>Flexible</td>
<td>Property taxation, urban planning</td>
<td>Health regulation</td>
<td>Includes phone books, mailing, credit reporting</td>
</tr>
<tr>
<td>Geographies of data collection</td>
<td>Flexible</td>
<td>Parcel</td>
<td>Business</td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td>Depends</td>
<td>Updated but not checked</td>
<td>Updated but not checked</td>
<td>Updated but not checked</td>
</tr>
<tr>
<td>Time of collection</td>
<td>Flexible</td>
<td>Varies</td>
<td>Typically updated annually</td>
<td>Updated regularly; some in time series</td>
</tr>
<tr>
<td>Cost—cleaning, coding</td>
<td>Expensive</td>
<td>Public data</td>
<td>Public data; may require significant formatting</td>
<td>Can be very expensive</td>
</tr>
</tbody>
</table>
3. Dilemmas: Data Quality
Accuracy and Completeness

• Address matching at first try and after extensive (weeks of) address cleaning and searching for Dun and Bradstreet data
• Business Analyst/Info USA that claims 90% matched to street address
• No one reports what they did

<table>
<thead>
<tr>
<th>Dataset (Twin Cities)</th>
<th>Number of Businesses</th>
<th>Unmatched First Try</th>
<th>Unmatched After Fixes (6 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eating and Drinking</td>
<td>4865</td>
<td>713 (16%)</td>
<td>97 (2%)</td>
</tr>
<tr>
<td>Food Stores</td>
<td>1806</td>
<td>286 (16%)</td>
<td>39 (2%)</td>
</tr>
<tr>
<td>General Department Stores</td>
<td>394</td>
<td>87 (22%)</td>
<td>22 (6%)</td>
</tr>
<tr>
<td>Liquor Stores</td>
<td>443</td>
<td>91 (21%)</td>
<td>2 (0%)</td>
</tr>
</tbody>
</table>

3. Dilemmas: Buffers
Network Buffer Change

• Network buffers indicate the area reachable within a certain street distance
• However, the formula inside ArcGIS changed with version 9.1 and a new Network Analyst
• The old style (ArcView 3.3) “joined the dots” of points x meters along the street network
• The new styles do not
• Can be 500%+ different in area

Take Home

• GIS software and databases are primarily developed for practice so don’t fit research questions
• Many researchers are unaware of weaknesses
• Most measurement problems can be solved with more research $\$
• But there’s still the issue of the built environment having small effects….and with so many measures some associations will be found by chance

There’s More than PA and Foods

• Designforhealth.net topics
  - Accessibility
  - Air quality
  - Climate change
  - Environment + housing quality
  - Food
  - Healthcare access
  - Mental health
  - Noise
  - Physical activity
  - Safety
  - Social capital
  - Water quality
  - Also special populations, planning process, finance
Built Environment, Food, and Physical Activity: Measurement Issues

1. Built Environment and Behavior—Measuring Relevant issues
2. GIS Measures
   - Basic approach, data, variables, and geographies
3. Typical Dilemmas
   - Definitions, data quality, buffering changes
4. Twin Cities Walking Study