

The Built Environment, Food, and Physical Activity: Measurement Issues

Ann Forsyth, PhD
Department of City and Regional Planning
Cornell University
January 19, 2009



Built Environment, Food, and Physical Activity: Measurement Issues

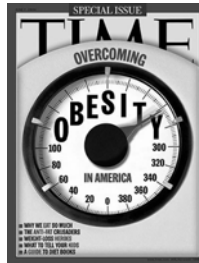


1. Built Environment and Behavior—Measuring Relevant issues
 2. GIS Measures
 - Basic approach, data, variables, and geographies
 3. Typical Dilemmas
 - Definitions, data quality, buffering changes
- My studies 2003-2006 TCWS (RWJF, PI) and cluster of NIH-funded studies from 2006 (IDEA, EAT-III, ECHO, 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. Built Environment

Why People Die

	US (CDC)	World (WHO)
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	4.4%	2.3% (#9, road)*
6. Diabetes	3.0%	
7. Flu & pneumonia	2.7%	6.9% (#3)
8. Alzheimer's/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)

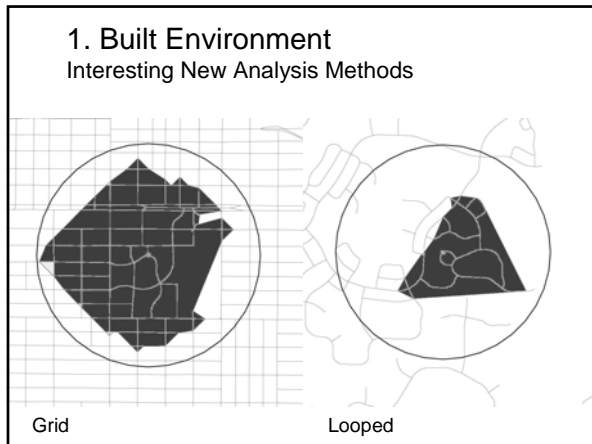


1. Built Environment

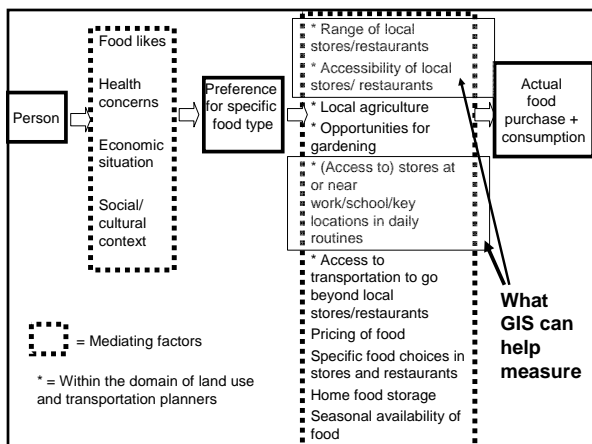
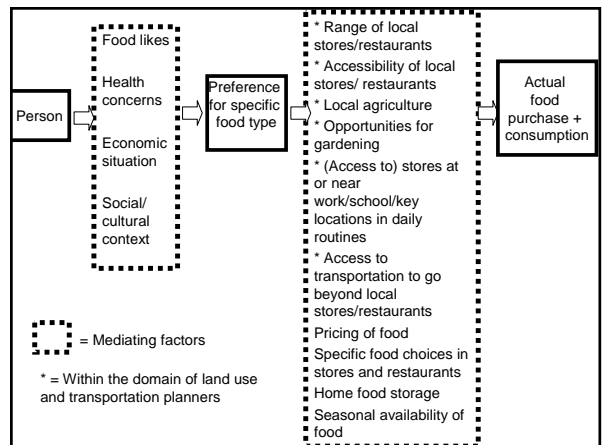
Theories about What Matters for Walking (and Biking)

- Transportation
- Urban/landscape design (physical planning, architecture, landscape architecture)
- Physical activity research





- ### 1. Built Environment
- #### 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 Colleegetown 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



- ### 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



<http://www.designforhealth.net/GISprotocols.html>



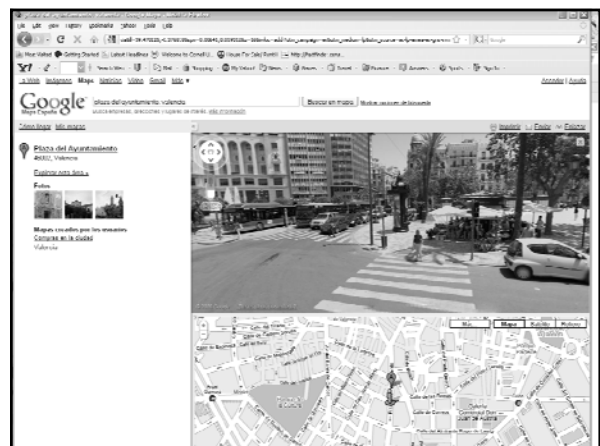
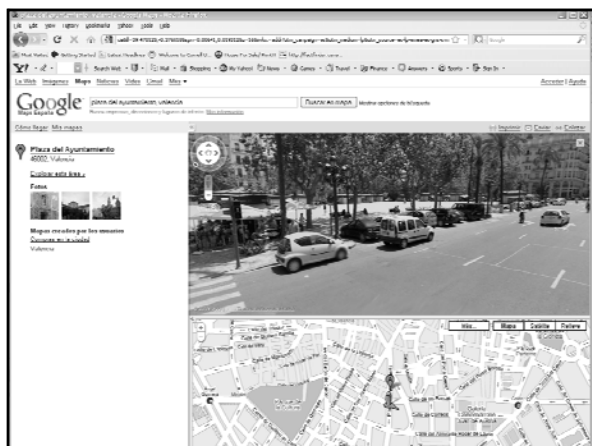
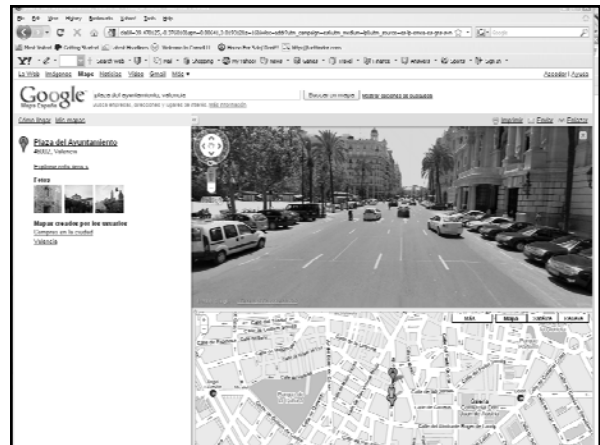
2. Measures How to Think About Measures

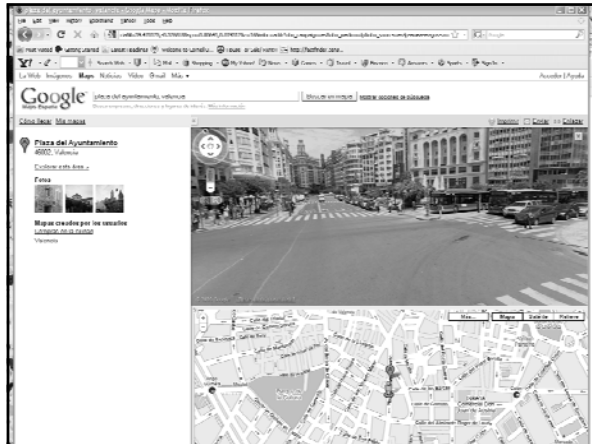
Measurement of food/activity variables associated with the built environment occurs on several levels:

- 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

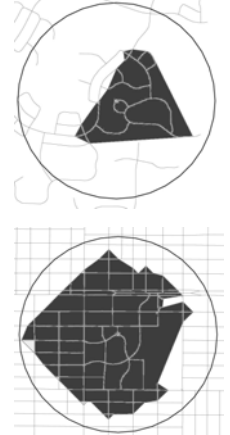
BE Variables In TREC/IDEA (45 + census)

- 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

2. Measures

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)



Aggregation or zoning effect



3. Dilemmas

Food Environments

- Definitions
- Data quality
- Buffering changes

3. Dilemmas: Definitions

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

Data Source (down)	Categorization				
	Researcher defined categories	Land use type	Chain names	Industrial code (e.g. NAICS)	Proprietary/ local codes
Fieldwork	xx		expensive		
Land use and parcel data		May not be specific	?		xx
Food licensing data			xx	?	?
Commercial business data			xx	xx	?

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

3. Dilemmas: Definitions

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
 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

3. Dilemmas: Definitions

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>

3. Dilemmas: Definitions

(Mis) Classification

Comparison of NAICS Code Distribution for 50 Zip Codes	Business Analyst	Dun and Brad.	Diff %
NAICS Code			
455 Food and Beverage Stores	452	394	-13%
4451 Grocery Stores	242	218	-10%
44511 Supermarkets and Other Grocery (except Convenience) Stores	117	70	-40%
4452 Specialty Food Stores	119	82	-31%
44521 Meat Markets	17	18	6%
44522 Fish and Seafood Markets	1	2	100%
44523 Fruit and Vegetable Markets	12	2	-83%
44529 Other Specialty Food Stores	88	59	-33%
445291 Baked Goods Stores	0	21	--
445292 Confectionery and Nut Stores	12	13	8%
445299 All Other Spec. Food Stores	76	26	-66%

3. Dilemmas: Definitions

(Mis)Classification Part 2

Comparison of NAICS Code Distribution for 50 Zip Codes	Business Analyst	Dun and Brad.	Diff %
722 Food Services and Drinking Places	1004	950	-5%
7221 Full-Service Restaurants	34	432	1171%
7222 Limited-Service Eating Places	829	371	-55%
722212 Cafeterias	5	8	60%
722213 Snack and Nonalcoholic Beverage Bars	75	1	-99%
7224 Drinking Places (Alcoholic Beverages)	91	103	13%

Comparing commercial numbers, correlation is high $r = .96$ 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

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

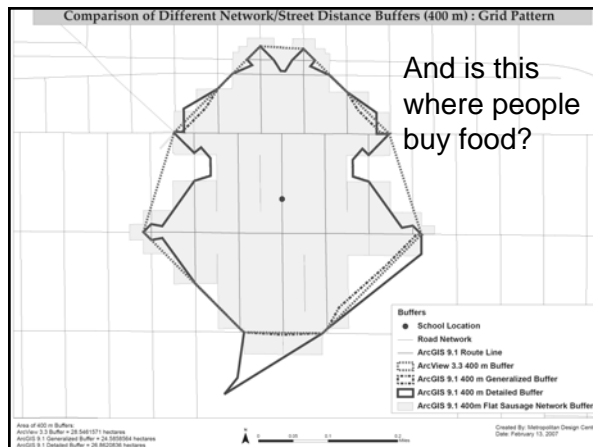
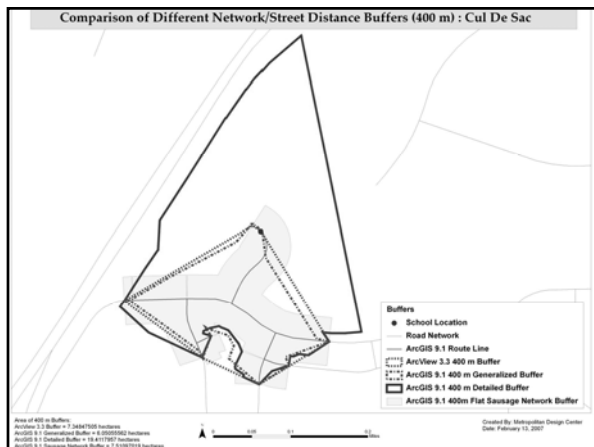
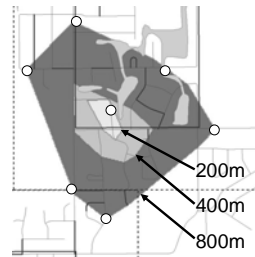
3. Dilemmas: Data Quality Accuracy and Completeness

- 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

Dataset (Twin Cities)	Number of Businesses	Unmatched First Try	Unmatched After Fixes (6 weeks)
Eating and Drinking	4565	713 (16%)	97 (2%)
Food Stores	1806	286 (16%)	39 (2%)
General Department Stores	394	87 (22%)	22 (6%)
Liquor Stores	443	91 (21%)	2 (0%)

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