ONEdata Obesity & Neighborhood Environment Database

Misa Graff, Ph.D.
On behalf of the ONEdata Group
University of North Carolina at Chapel Hill

ONEdata
Obesity & Neighborhood Environment database

Purpose of ONEdata: Characterize the Obesogenic Neighborhood

ONEdata
Obesity & Neighborhood Environment database

Obesity & Environment

- Research findings
- Database development process
- Overview of environment data

ONEdata
Obesity & Neighborhood Environment database

Obesity and Environment Research Findings

ONEdata
Obesity & Neighborhood Environment database

Background

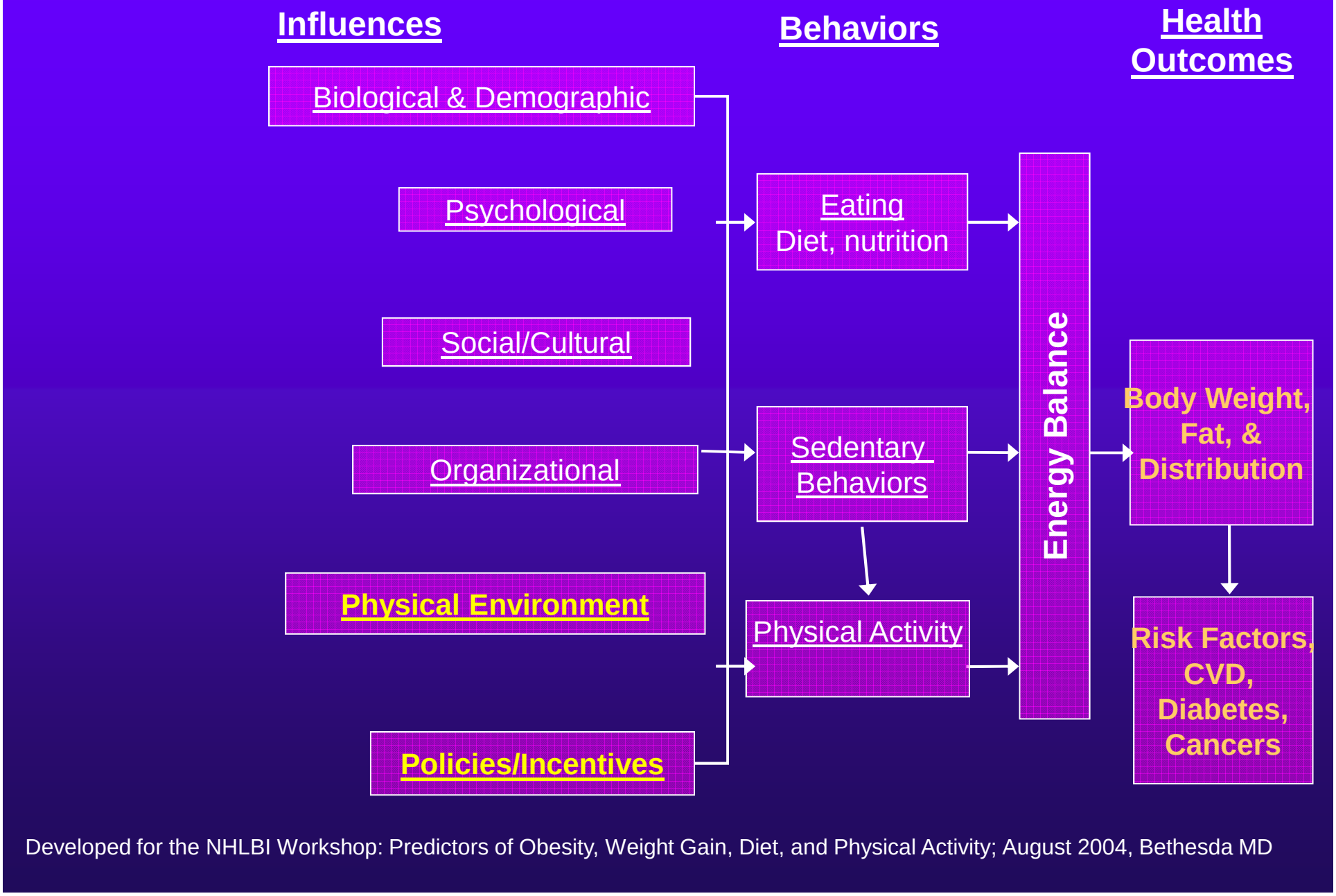
- Adolescence and young adulthood are major periods of biological, social and behavioral development, with potential importance for future health behaviors.

- 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**

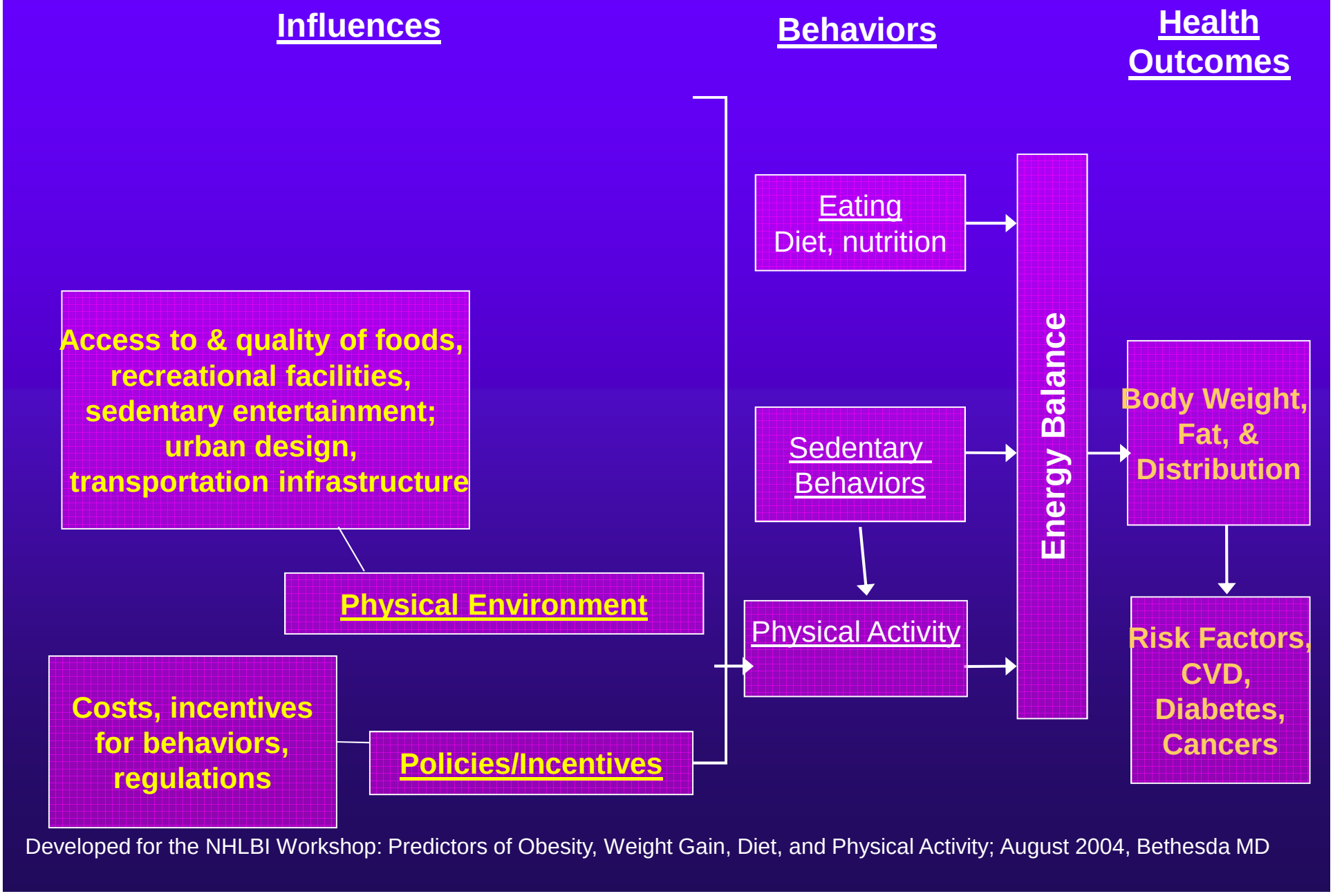
ONEdata
Obesity & Neighborhood Environment database

An Ecological Model of Diet, Physical Activity, and Obesity



Developed for the NHLBI Workshop: Predictors of Obesity, Weight Gain, Diet, and Physical Activity; August 2004, Bethesda MD

An Ecological Model of Diet, Physical Activity, and Obesity



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

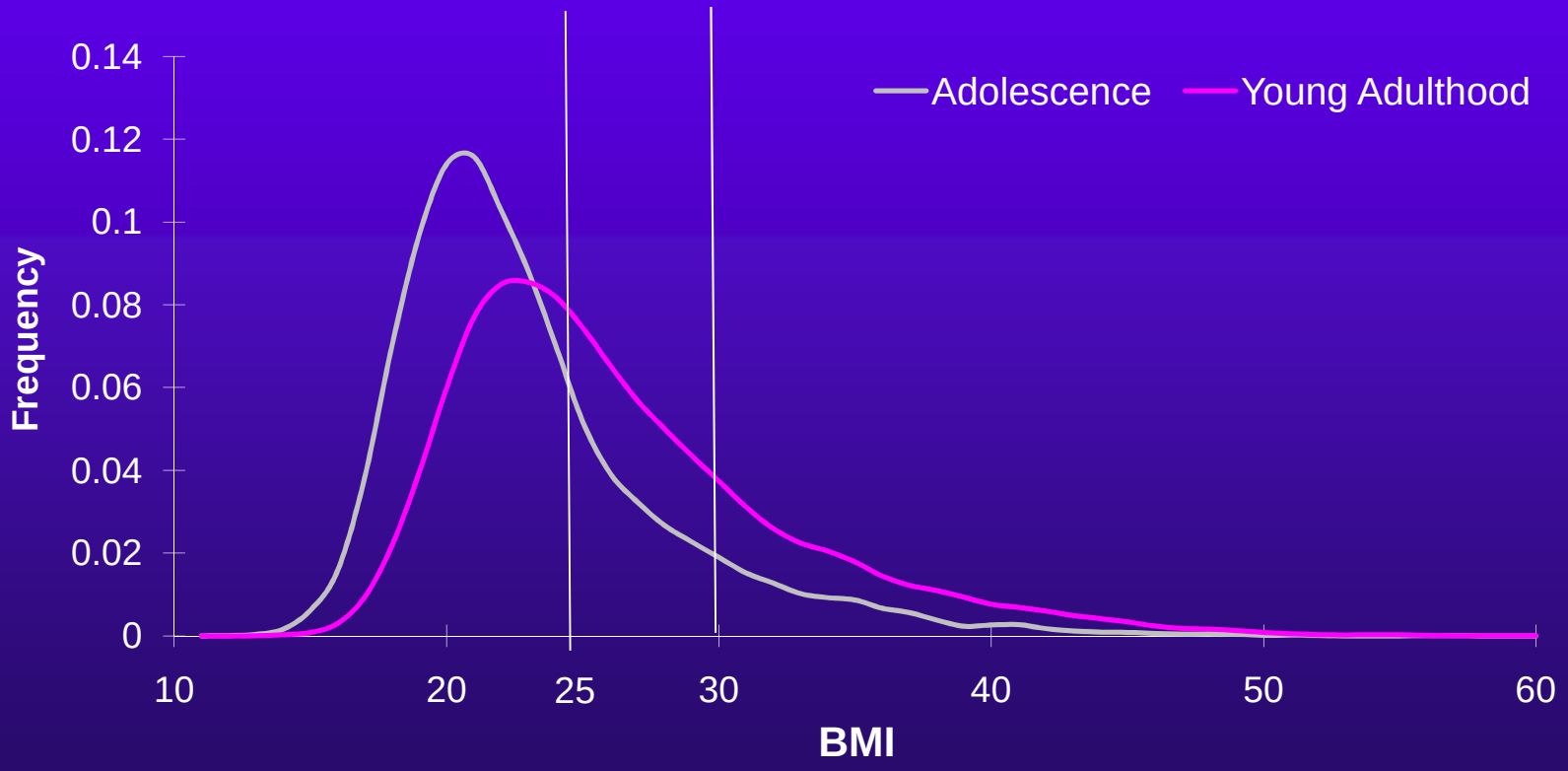
ONEdata
Obesity & Neighborhood Environment database

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

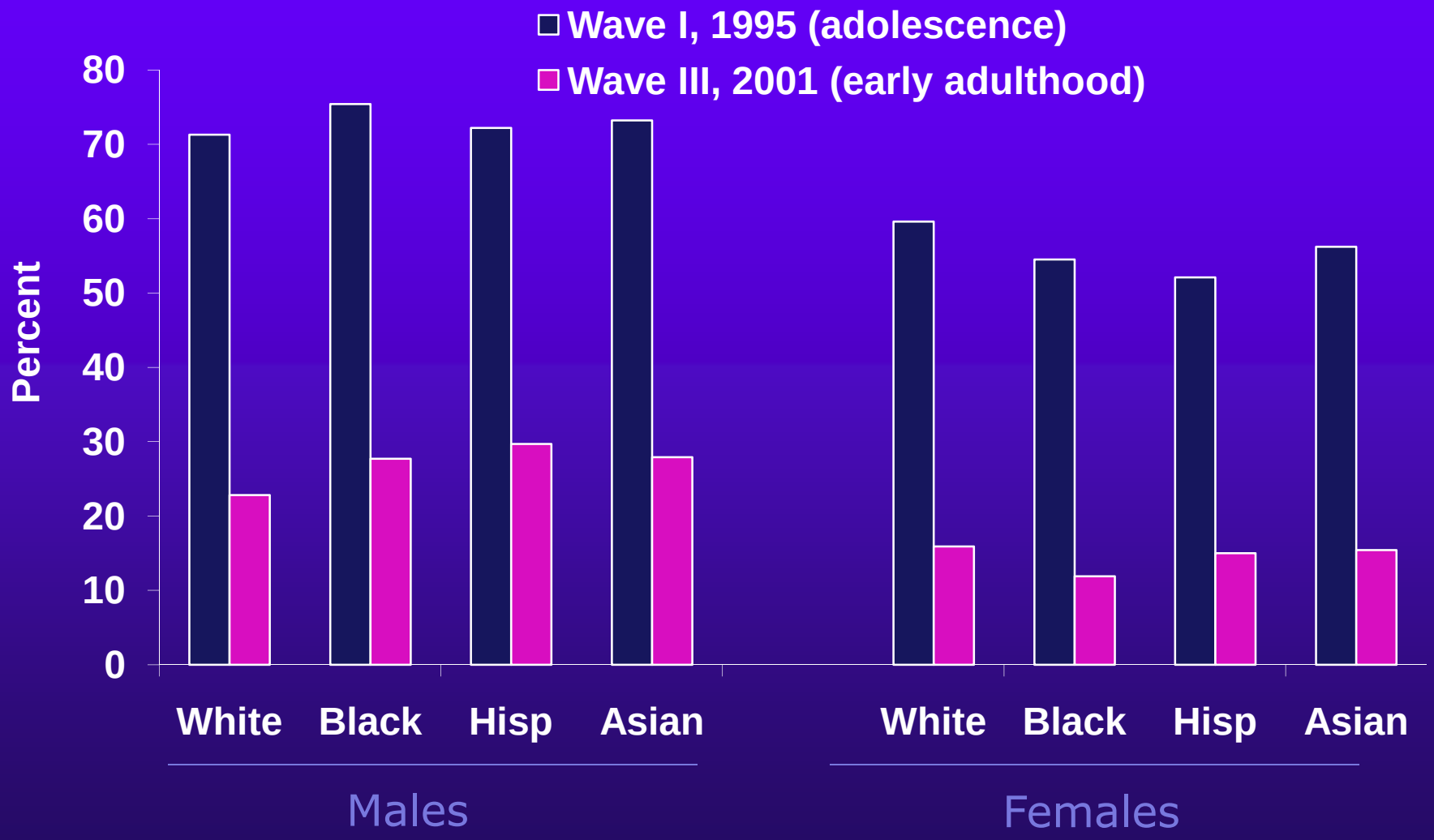
ONEdata
Obesity & Neighborhood Environment database

BMI Distribution of Add Health respondents at adolescence and young adulthood



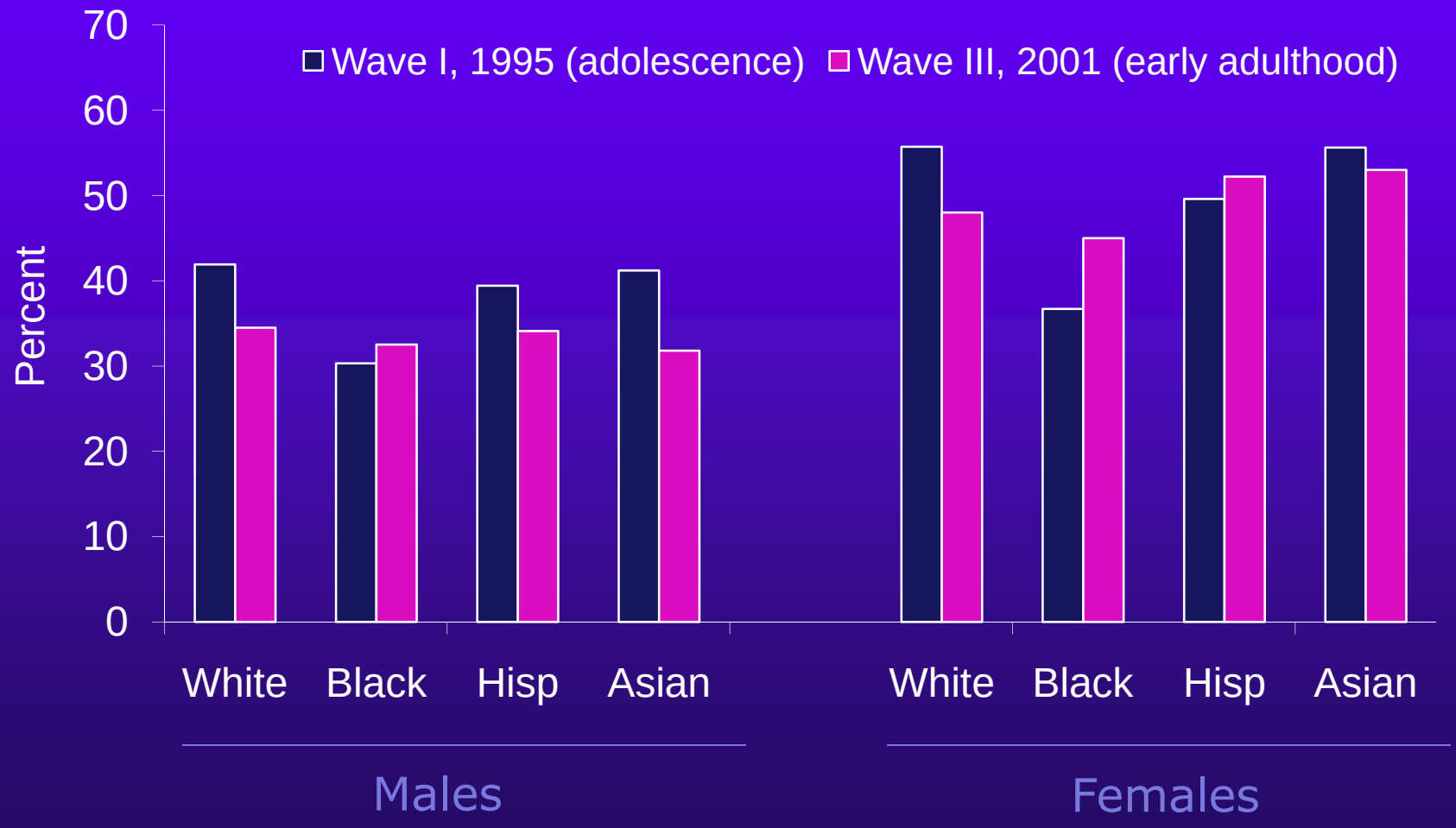
Gordon-Larsen, P. et al. 2004. Am J Clin Nutr

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



Gordon-Larsen, et al. Am J Prev Med 2004.

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



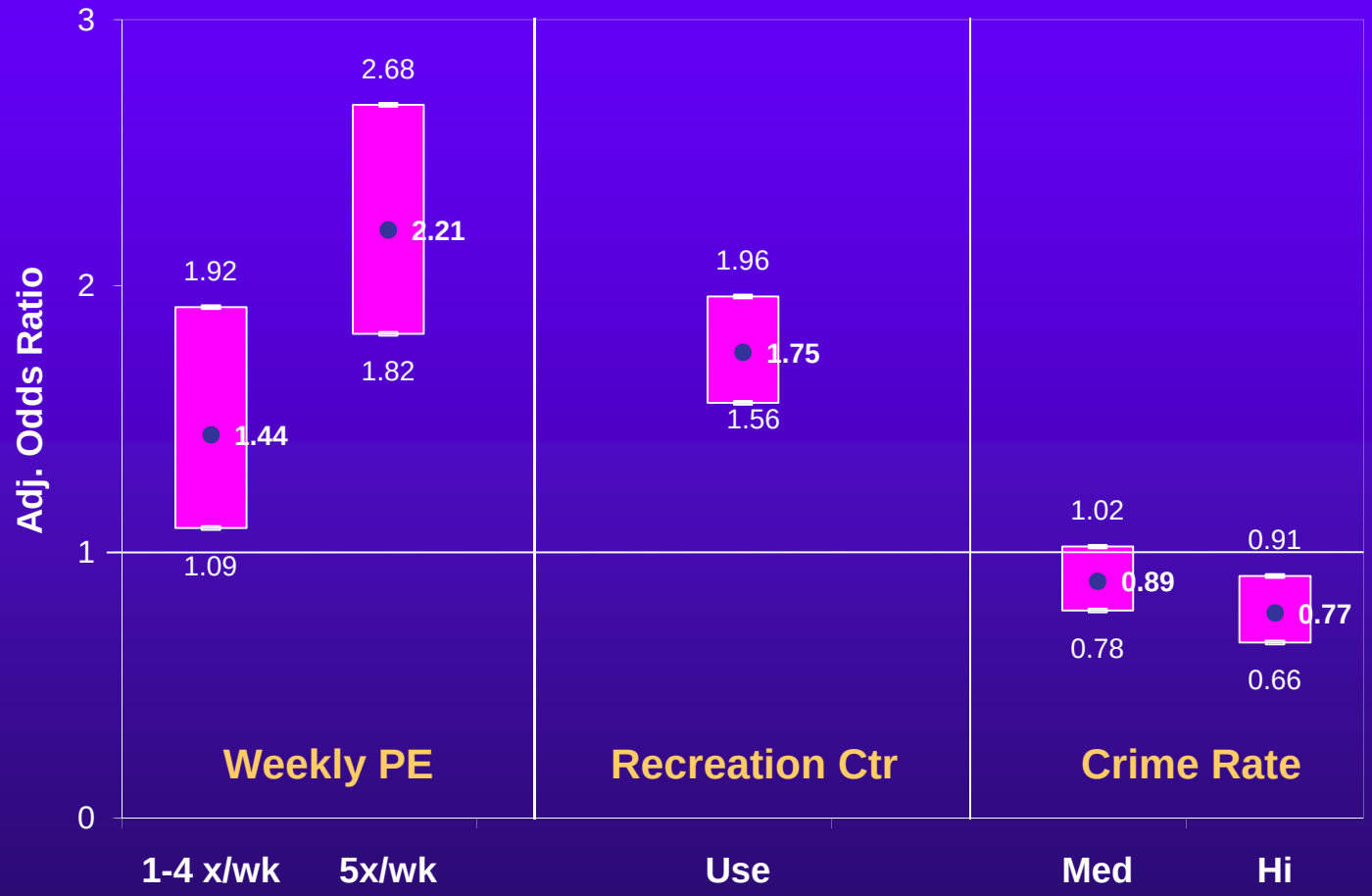
Gordon-Larsen, et al. Am J Prev Med 2004.

Research Question

- What about modifiable environmental factors?

ONEdata
Obesity & Neighborhood Environment database

There Are Important Associations Between Modifiable Environmental Factors and Activity Patterns of US Adolescents



Reference
no weekly PE
no rec ctr. use
ref=low crime

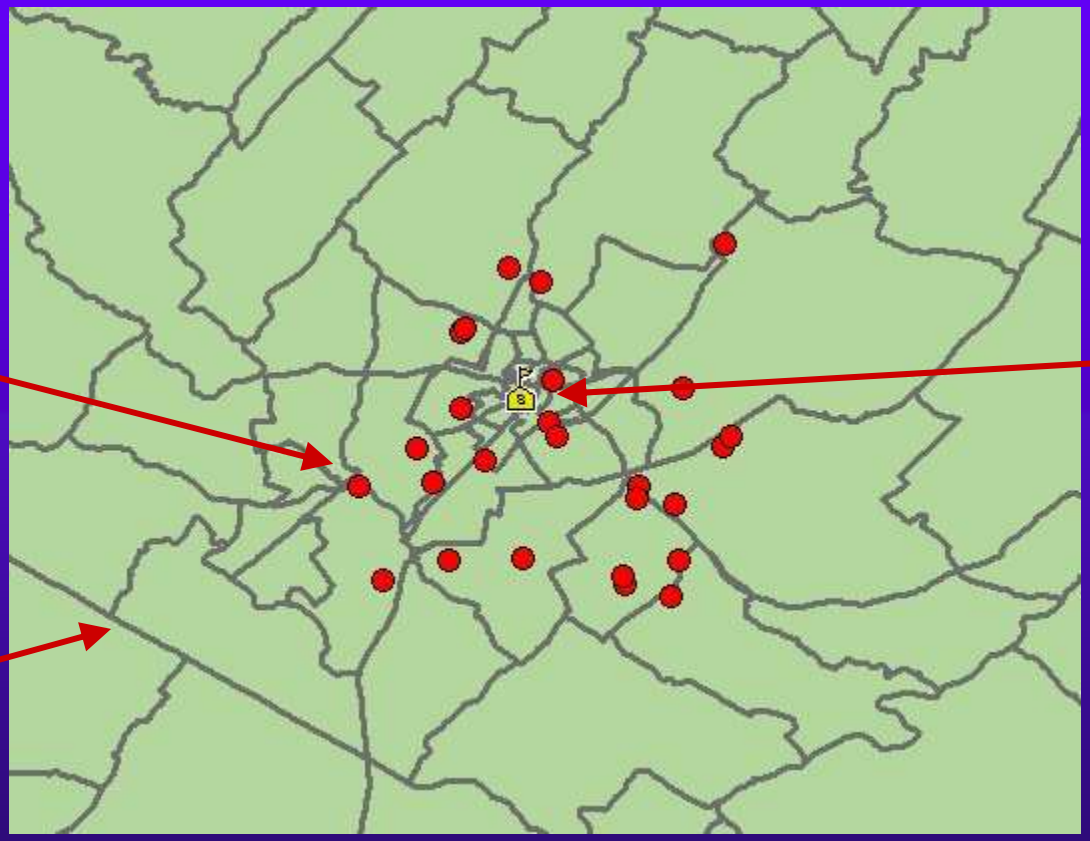
Adjusted for sex, age, SES, urban residence,
in-school status, pregnancy, region and month of interview

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



Add Health Neighborhoods
42,857 block groups w/in 5 mi of respondent

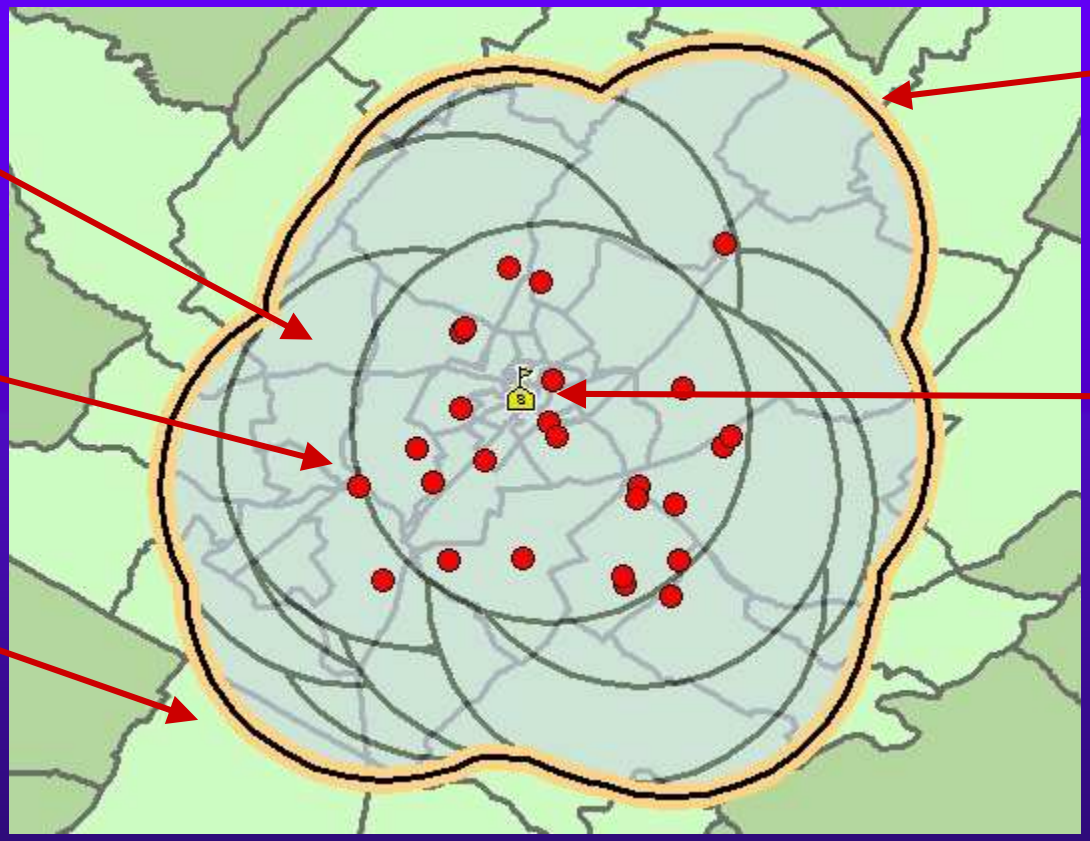
Individual Buffer

Community Study Area

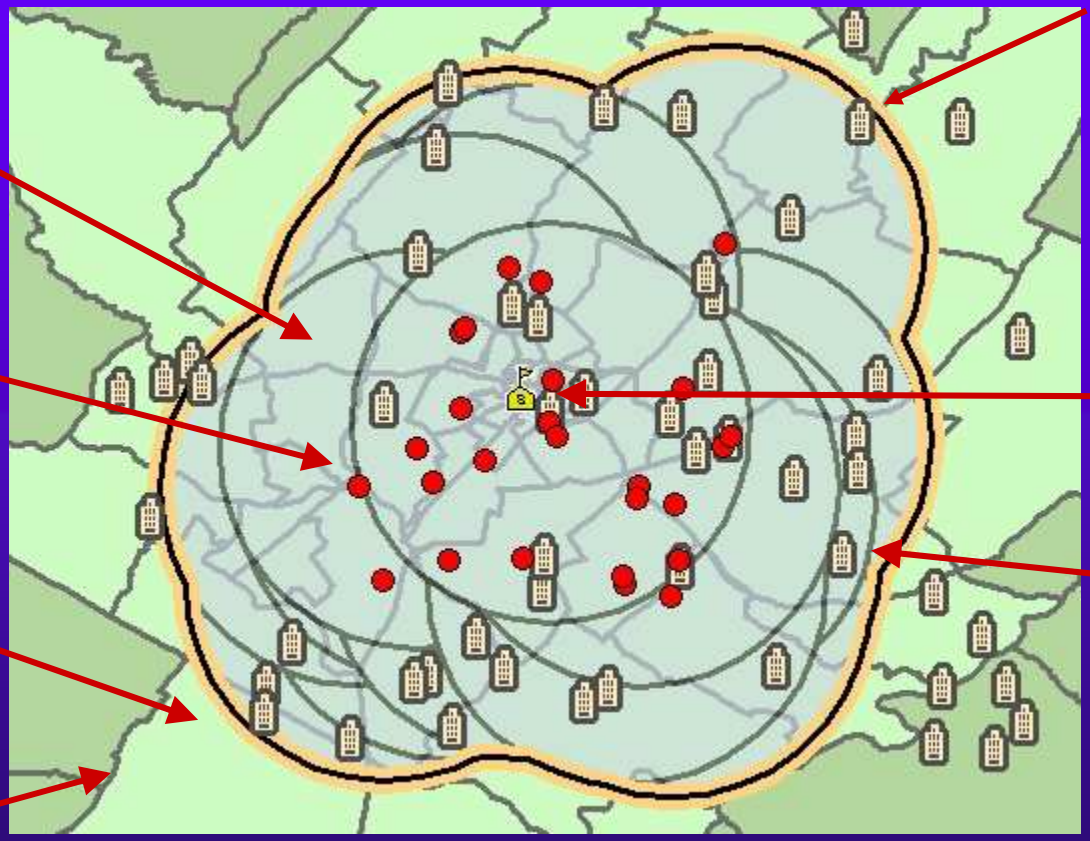
Respondents

School

Sampled Block Group



Add Health Neighborhoods
42,857 block groups w/in 5 mi of respondent



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

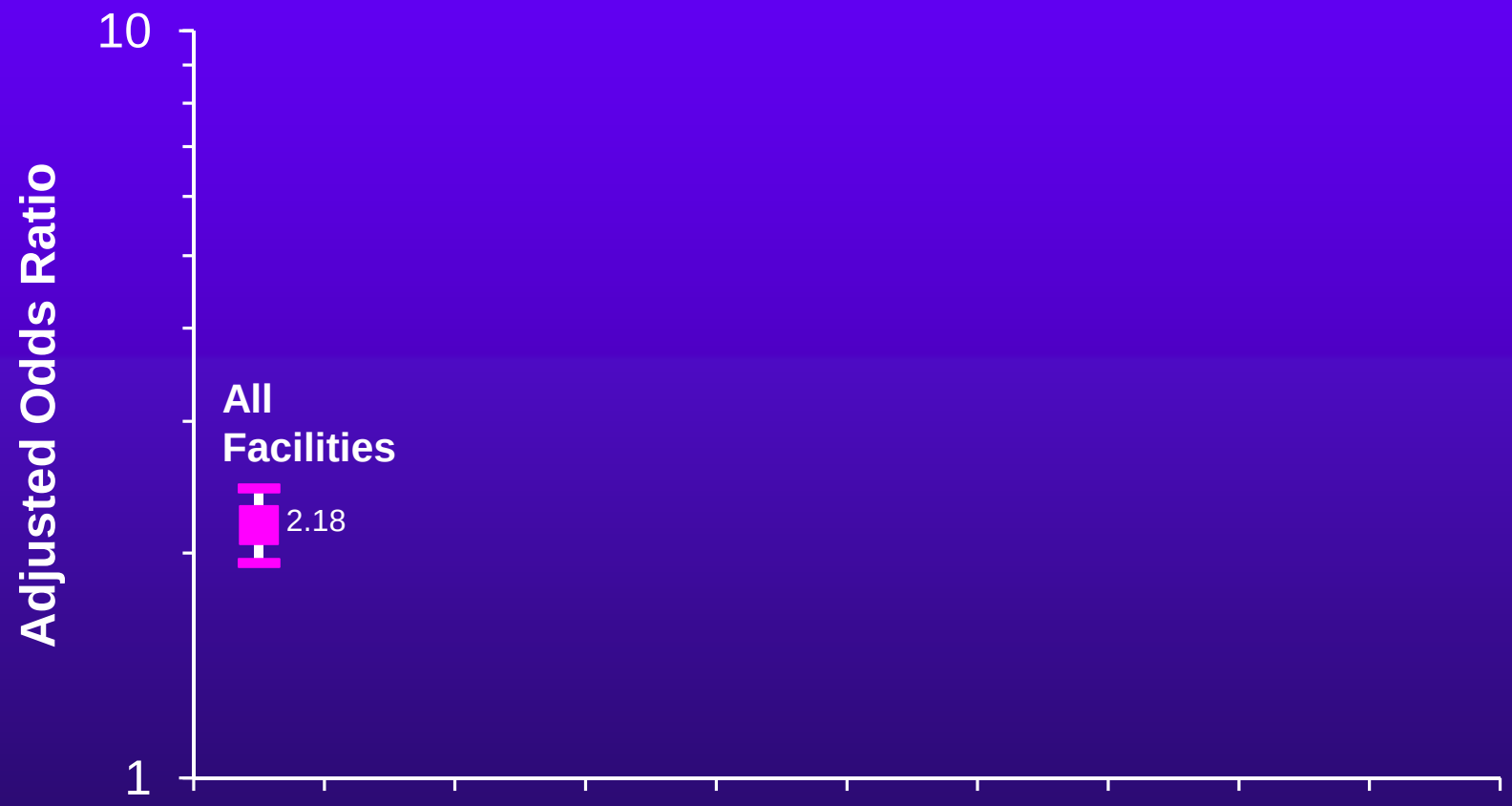
Analysis Methods

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‡

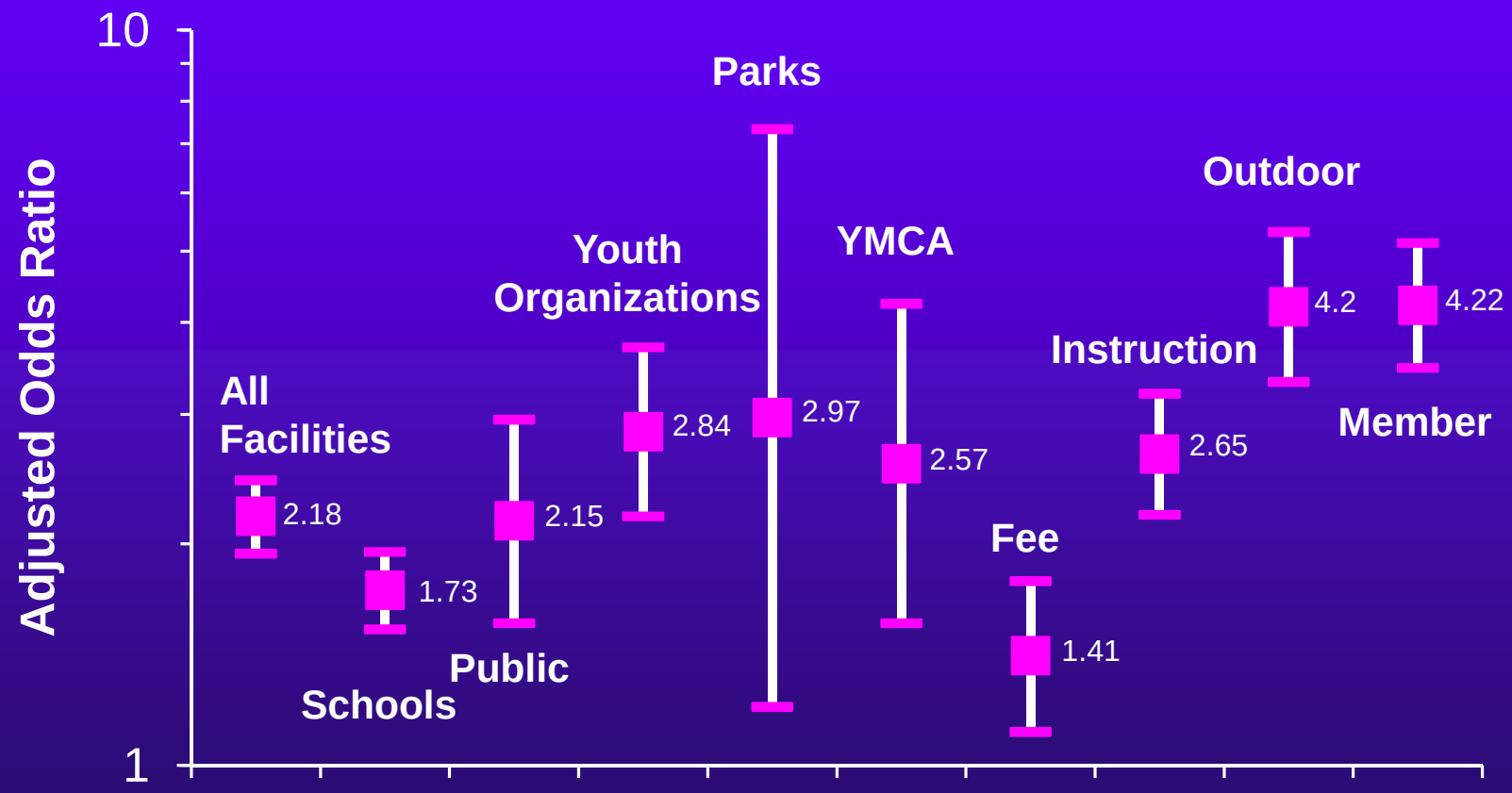


† Increased odds given each 100% increase in population with college education+

‡ Adjusted by population density and % minority population

Gordon-Larsen, et al. *Pediatrics* 2006

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

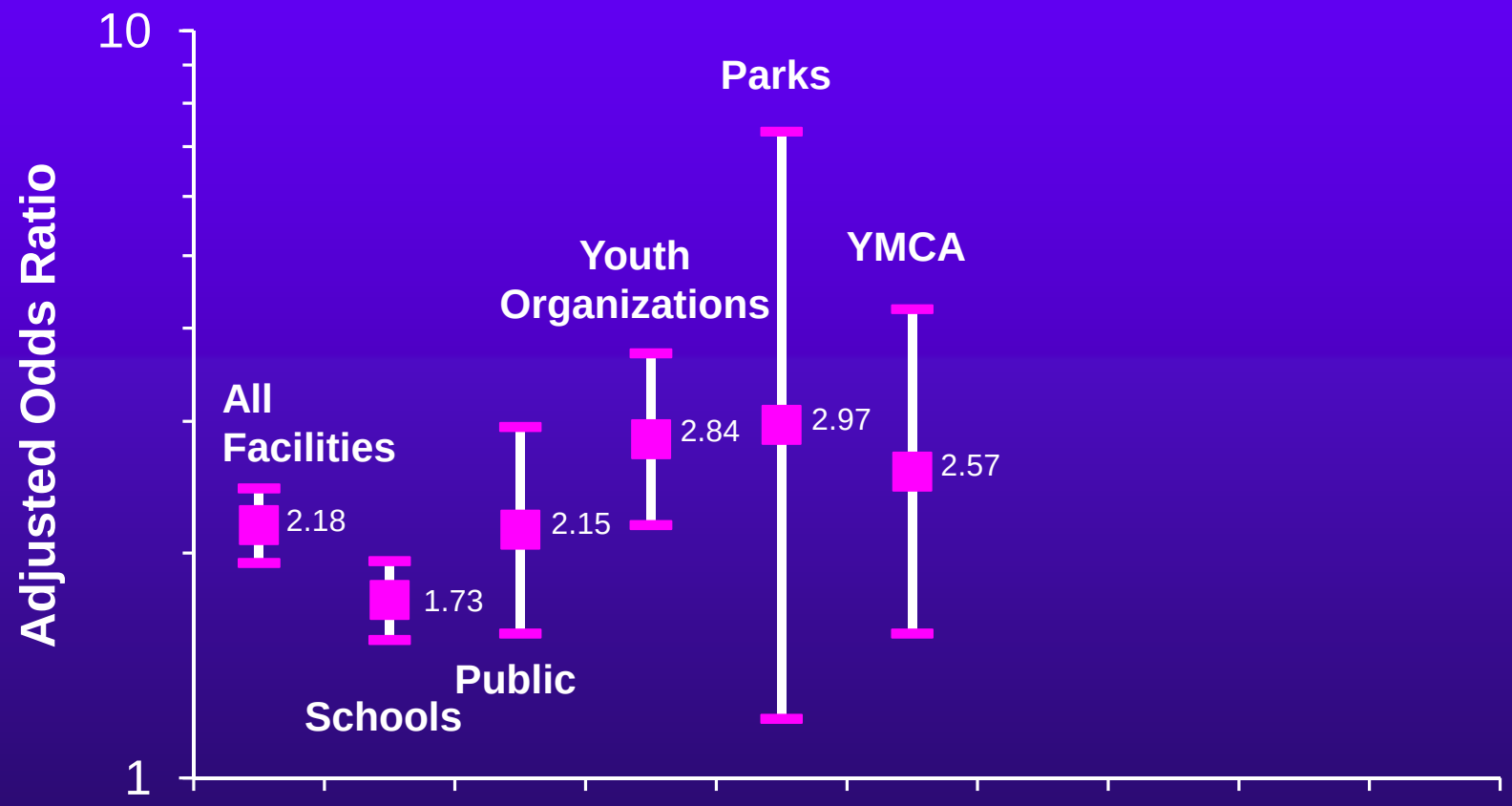


† Increased odds given each 100% increase in population with college education+

‡ Adjusted by population density and % minority population

Gordon-Larsen, et al. *Pediatrics* 2006

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

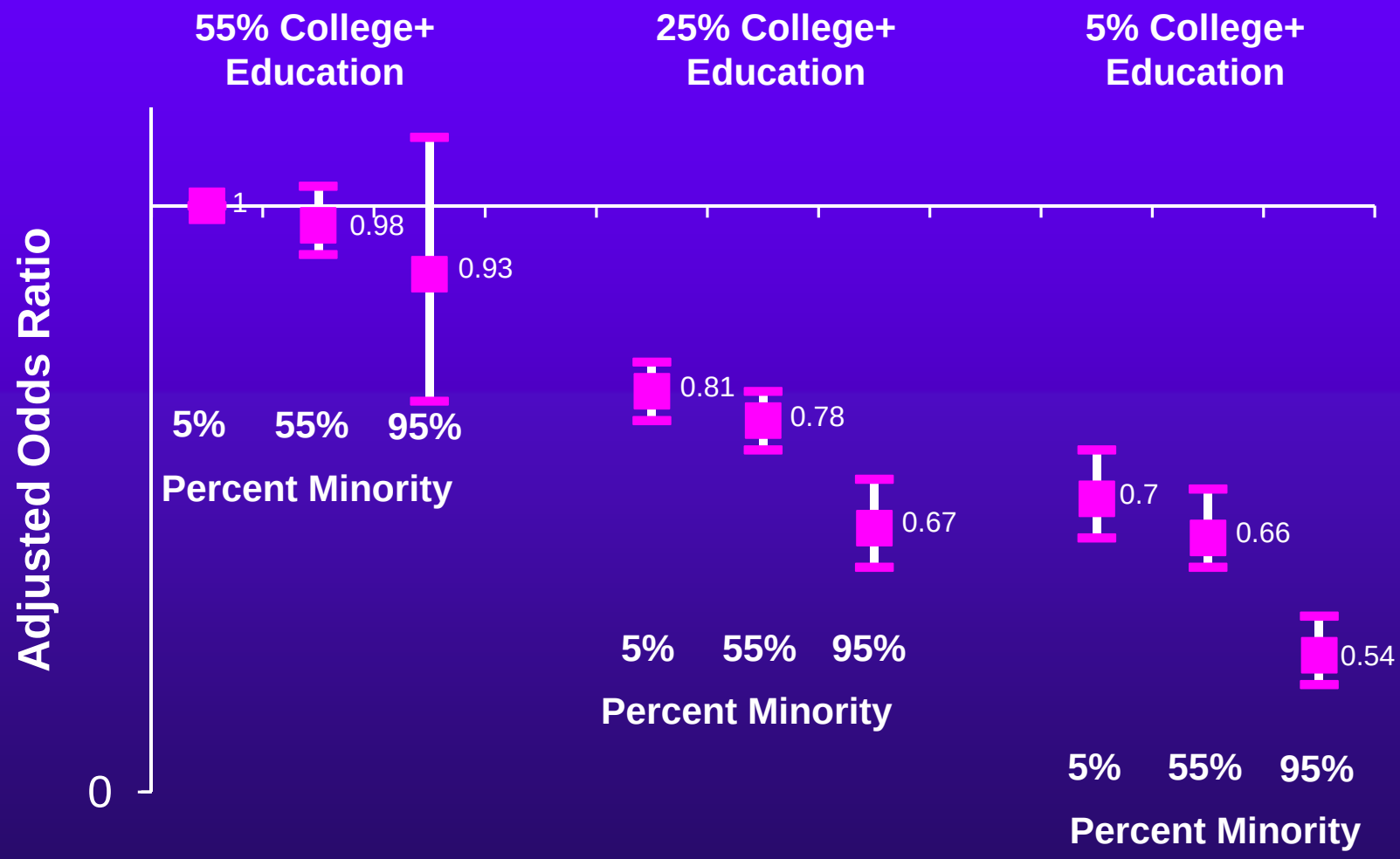


† Increased odds given each 100% increase in population with college education+

‡ Adjusted by population density and % minority population

Gordon-Larsen, et al. *Pediatrics* 2006

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

Gordon-Larsen, et al. *Pediatrics* 2006

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

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

Odds Ratios (95% CI) adjusting for population density

Gordon-Larsen, et al. *Pediatrics* 2006

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

ONEdata
Obesity & Neighborhood Environment database

Obesity and Environment Database Development Process

ONEdata
Obesity & Neighborhood Environment database

The Obesogenic Environment

Obesogenic environment

ONEdata
Obesity & Neighborhood Environment database

Obesogenic environment

ONEdata
Obesity & Neighborhood Environment database

Dan Burden, Walkable Communities Inc.

Obesogenic environment

ONEdata
Obesity &
Neighborhood
Environment
database

Dan Burden, Walkable Communities Inc.

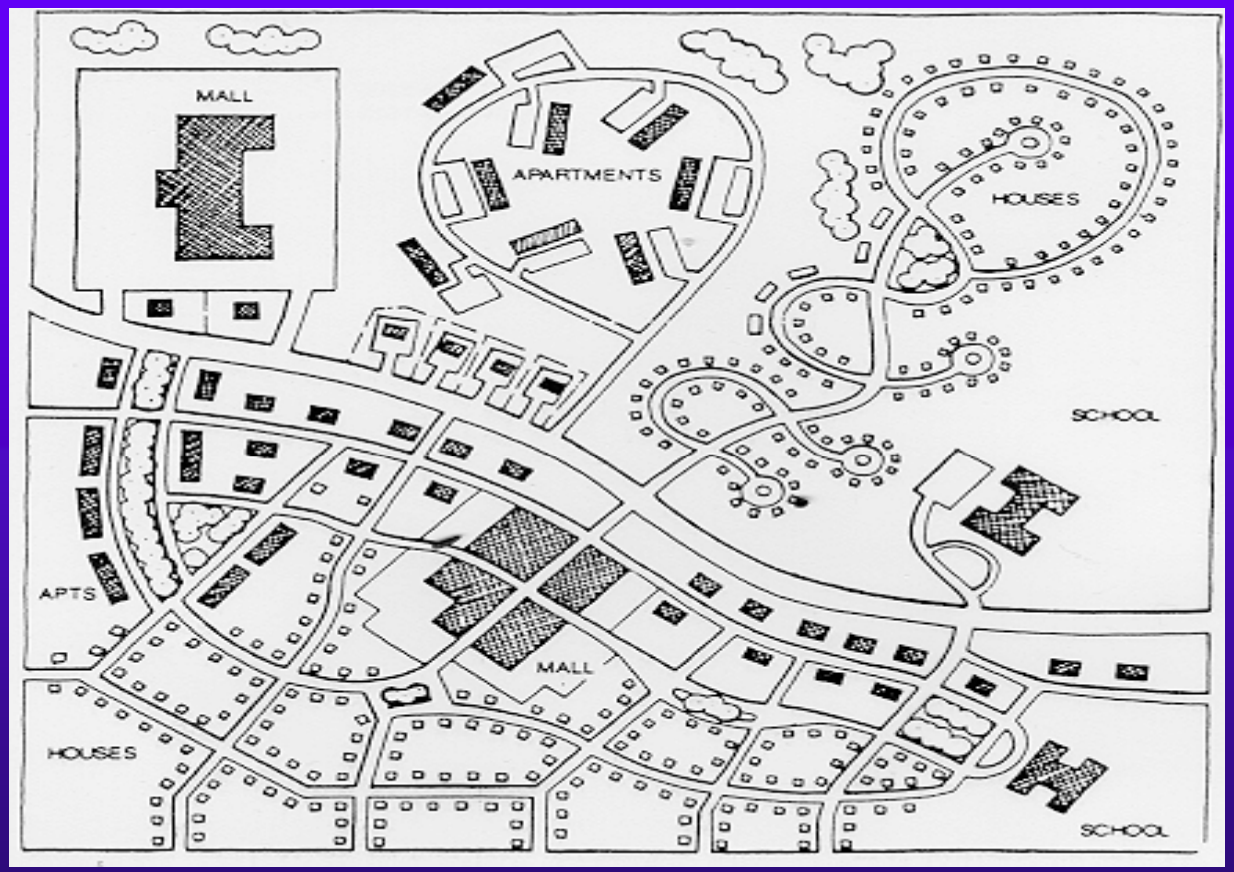
Obesogenic environment

ONEdata
Obesity & Neighborhood Environment database

The Built 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

ONEdata
Obesity & Neighborhood Environment database

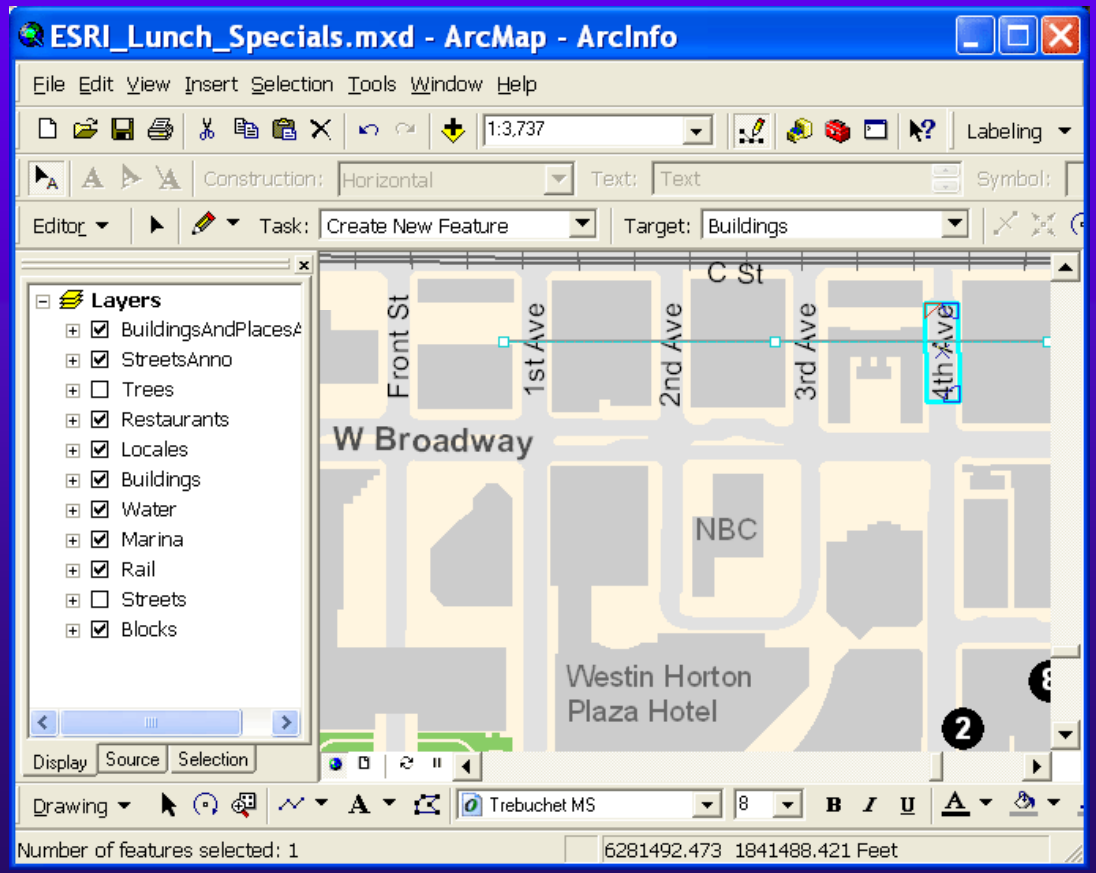
Definition of Built Environment

- "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"

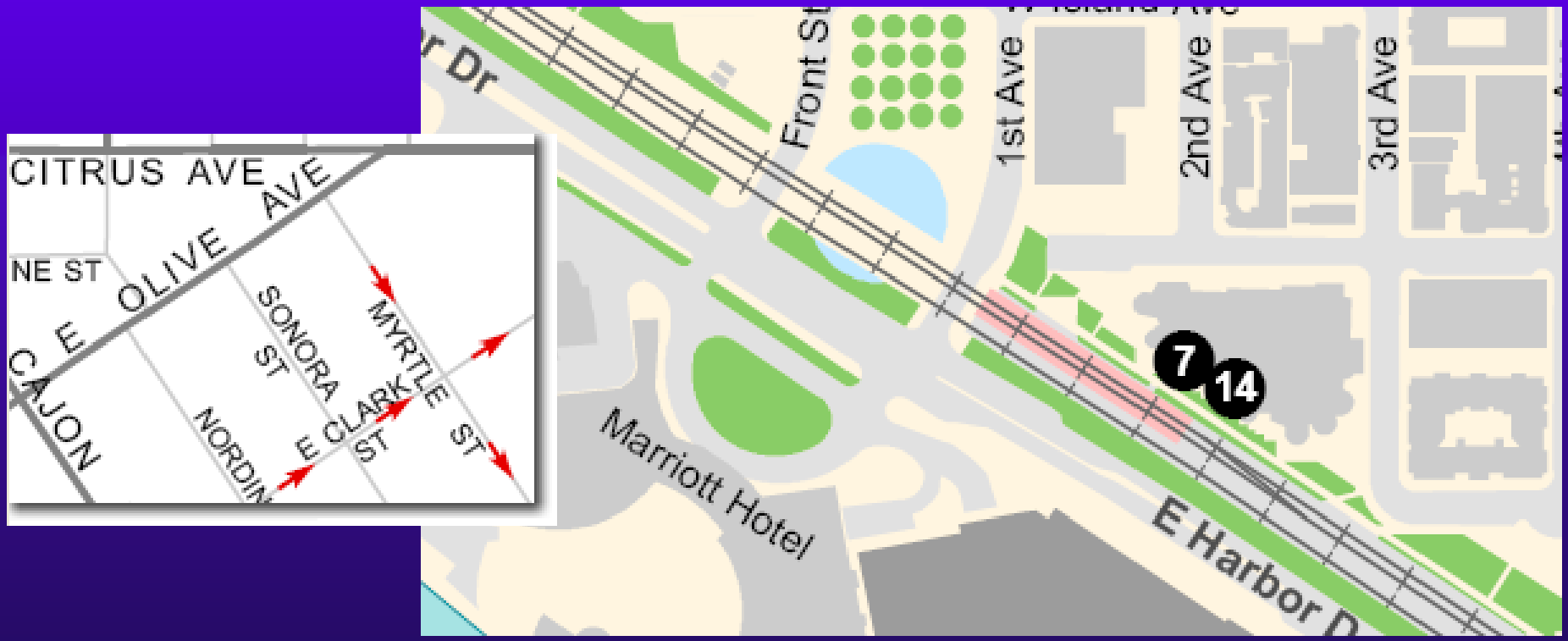
ONEdata
Obesity & Neighborhood Environment database

(RFA-ES-04-003, Sallis & Glanz 2006)

ESRI Mapping Center



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.

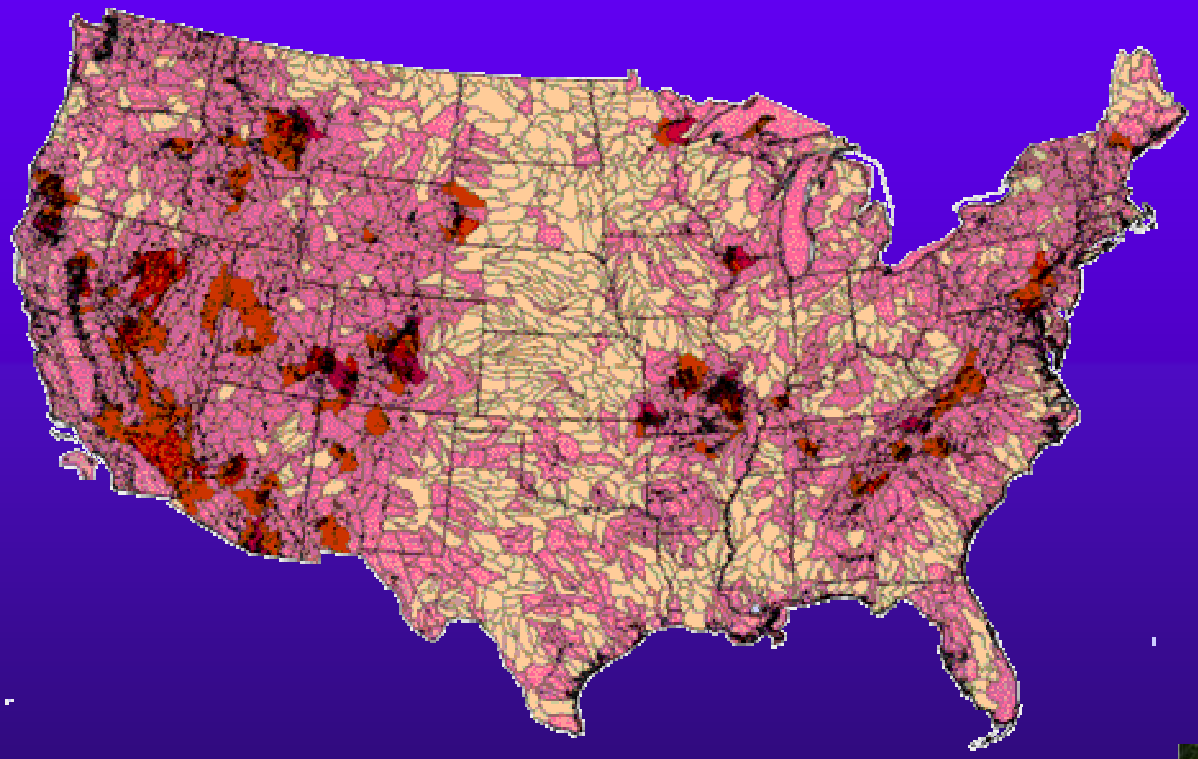
ONEdata
Obesity & Neighborhood Environment database

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



ONEdata
Obesity &
Neighborhood
Environment
database

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

ONEdata
Obesity & Neighborhood Environment database

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

ONEdata
Obesity & Neighborhood Environment database

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++)

ONEdata
Obesity & Neighborhood Environment database

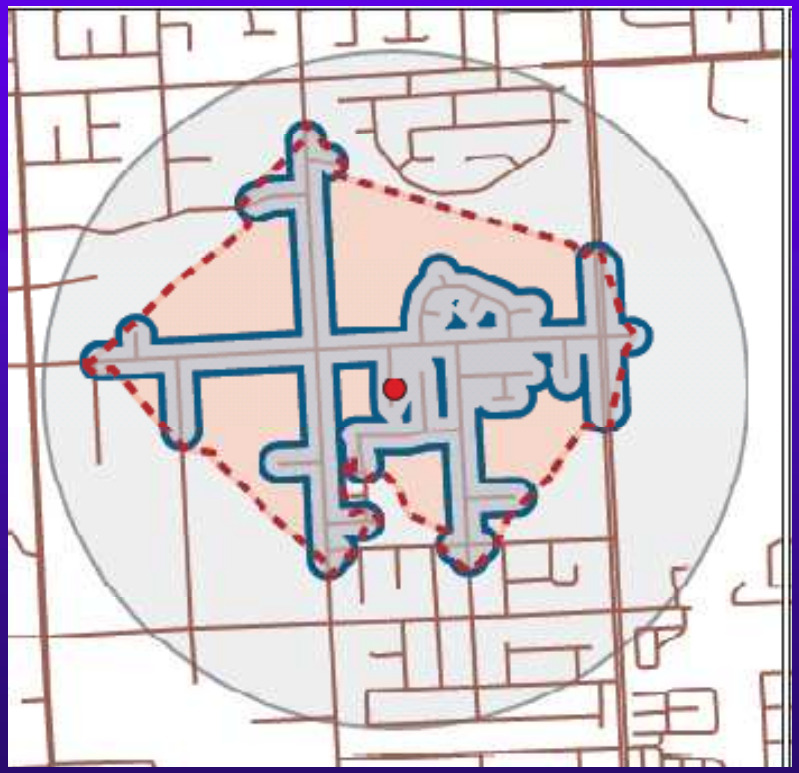
Obesity and Environment
Overview of Environment Data

ONEdata
Obesity & Neighborhood Environment database

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

- 145 communities, 80 counties, 33 states, >42,000 block groups
- 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

ONEdata
Obesity & Neighborhood Environment database

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)

ONEdata
Obesity & Neighborhood Environment database

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

Respondent Locations
Facilities

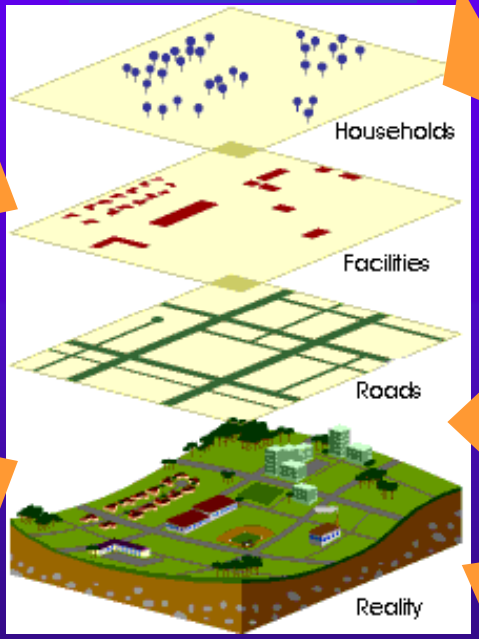
Contextual Data

U.S. Department of Labor
Bureau of Labor Statistics

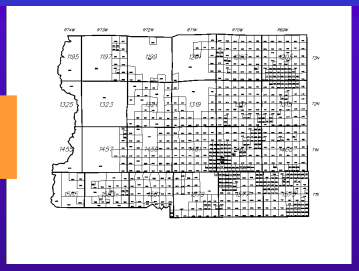
NOAA

United States Census 2000

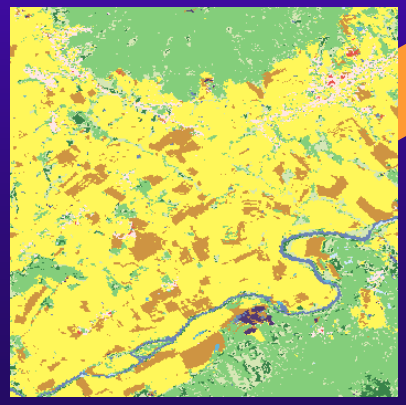
GIS Database



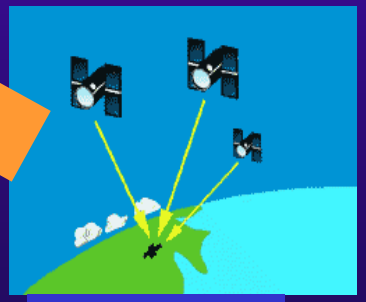
Ancillary Data
(Roads, Administrative Boundaries, etc.)



Land Use



Derived Measures
Distance Matrices, Network Calculations, Connectivity, Community Classifications



GPS Data

Join database to individual-level data

ID	BMI	BG: % in poverty	Number of Parks	Number of McDonalds
10090	31.5	43.2	0	10
10091	25.5	23.4	5	5
10000	28.9	19.1	8	0
10123	38.2	2.5	2	3

Some validation work

- Physical activity facilities database

Characteristics of the validation locations in two non-Add Health communities

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

†unweighted average among block groups, using 2000 Census data

Example of potential influence of GIS error on counts of facilities

Boone et al. Ann Epidemiol 2008

Agreement* of *administratively defined* neighborhood and nearest street of GPS and geocoded physical activity facility locations

Community	# Facilities	5-Digit ZIP Code	Census Tract	Block Group	Street
Non-Urban	63	59	59	58	45
		94%	94%	92%	71%
Urban	42	42	42	41	30
		100%	100%	98%	71%

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

Boone et al. Ann Epidemiol 2008

Challenges

Dataset Integration

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

ONEdata
Obesity & Neighborhood Environment database

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 are within in underlying street data used for respondent & other locations in one area

  → required spatial adjustment in GIS database.

Challenges

National data development effort → "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

ONEdata
Obesity & Neighborhood Environment database

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

ONEdata
Obesity & Neighborhood Environment database

For More Information

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

ONEdata
Obesity & Neighborhood Environment database