Using Multiple Data Sources in GIS for Public Health Research

Craig T. Dearfield

Howard University/AFYA Inc./Akeso Consulting

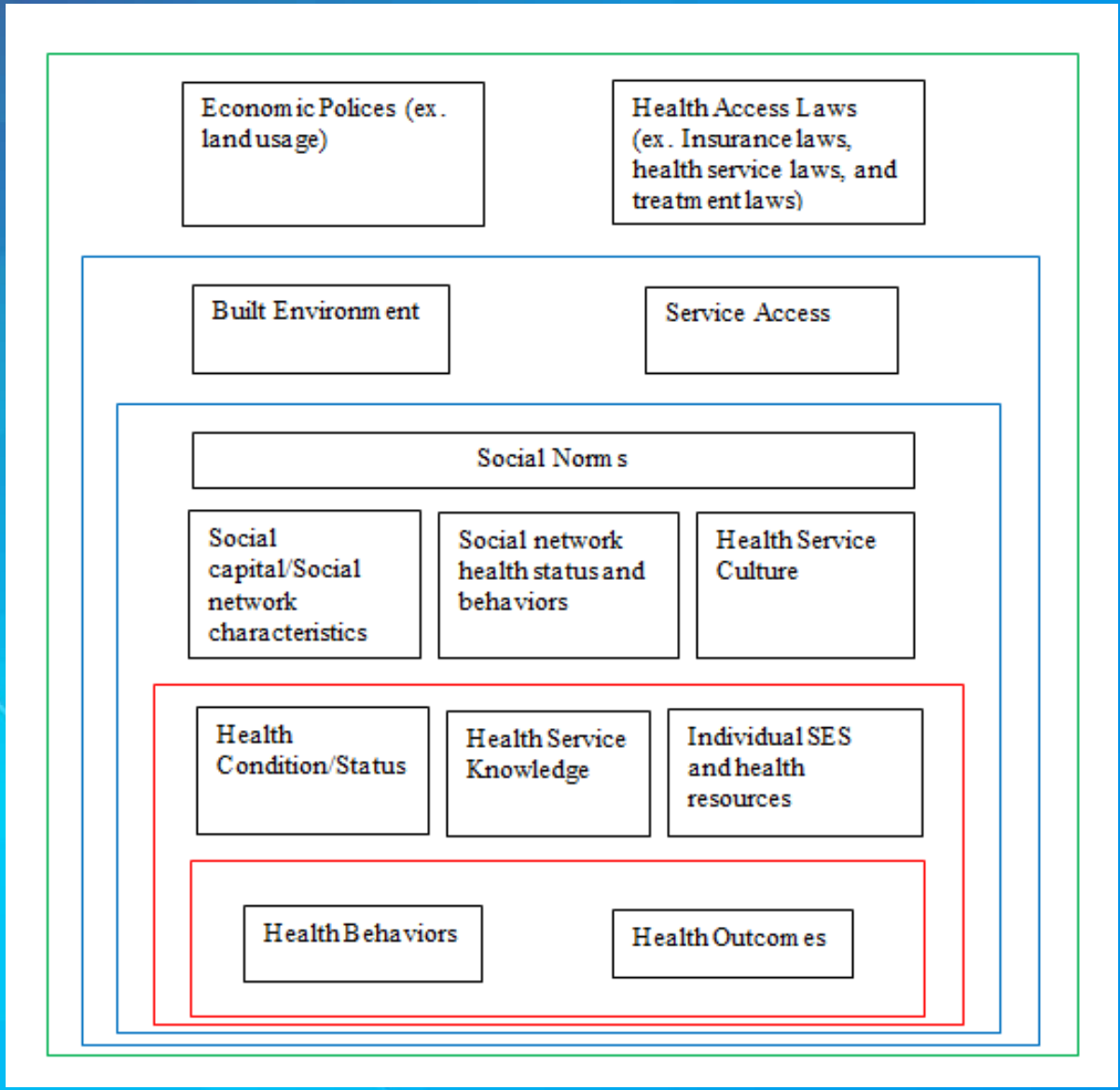
Introduction

- Health status, morbidity, and mortality are influenced by a wide range of individual and contextual factors.
- Individual factors typically include:
  - Biologic characteristics
  - Socioeconomic status
  - Socially defined categories such as race and ethnicity, and knowledge, beliefs, and behaviors.
- Contextual factors typically include:
  - Social networks
  - Neighborhood and environmental context
  - Political and economic influences

Introduction

- These factors can be understood as multiple, nested analytic levels.
- It is important to consider and measure factors at multiple analytic levels in public health research.

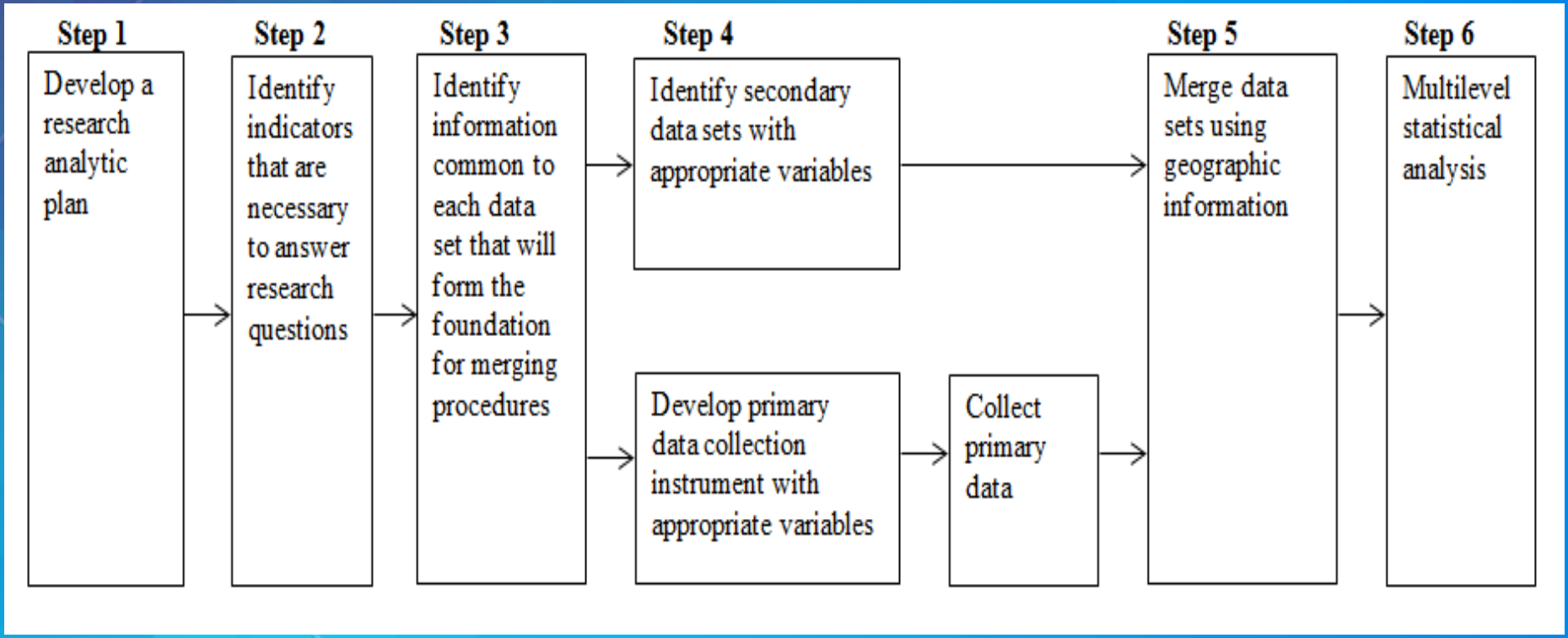
Introduction



Introduction

- Researchers may not have the resources needed to collect the data necessary to measure these variables at each analytic level as part of a single research project.
- Using secondary data sources can increase the scope of analysis beyond individual variables to include assessments of structural influences.

Methods


Step 1
Develop a research analytic plan
Step 2
Identify indicators that are necessary to answer research questions
Step 3
Identify information common to each data set that will form the foundation for merging procedures
Step 4
Identify secondary data sets with appropriate variables
Develop primary data collection instrument with appropriate variables
Collect primary data
Step 5
Merge data sets using geographic information
Step 6
Multilevel statistical analysis

Methods – Primary Data Collection

- Individual surveys in primary care offices in Washington, DC (Howard University Hospital)
- Individual-level attributes:
  - Demographics
  - Geographic location
  - Health status
  - Tobacco use and tobacco cessation behavior
- Respondent social support and social capital focusing on health-related topics, health resources, and support for positive health behaviors

Methods – Primary Data Collection

Characteristic	Value	Frequency
Gender	Male	54 (35.5%)
	Female	78 (51.3%)
Race/Ethnicity	African-American/Black	120 (78.4%)
	Asian	2 (1/3%)
	Native American/American Indian/Alaska Native	1 (.7%)
	Caucasian/White	1 (.7%)
	Other (mulitracial and/or ethnicity/nationality)	5 (3.3%)
Occupational Status	Employed	60 (39.5%)
	Unemployed	92 (60.5%)

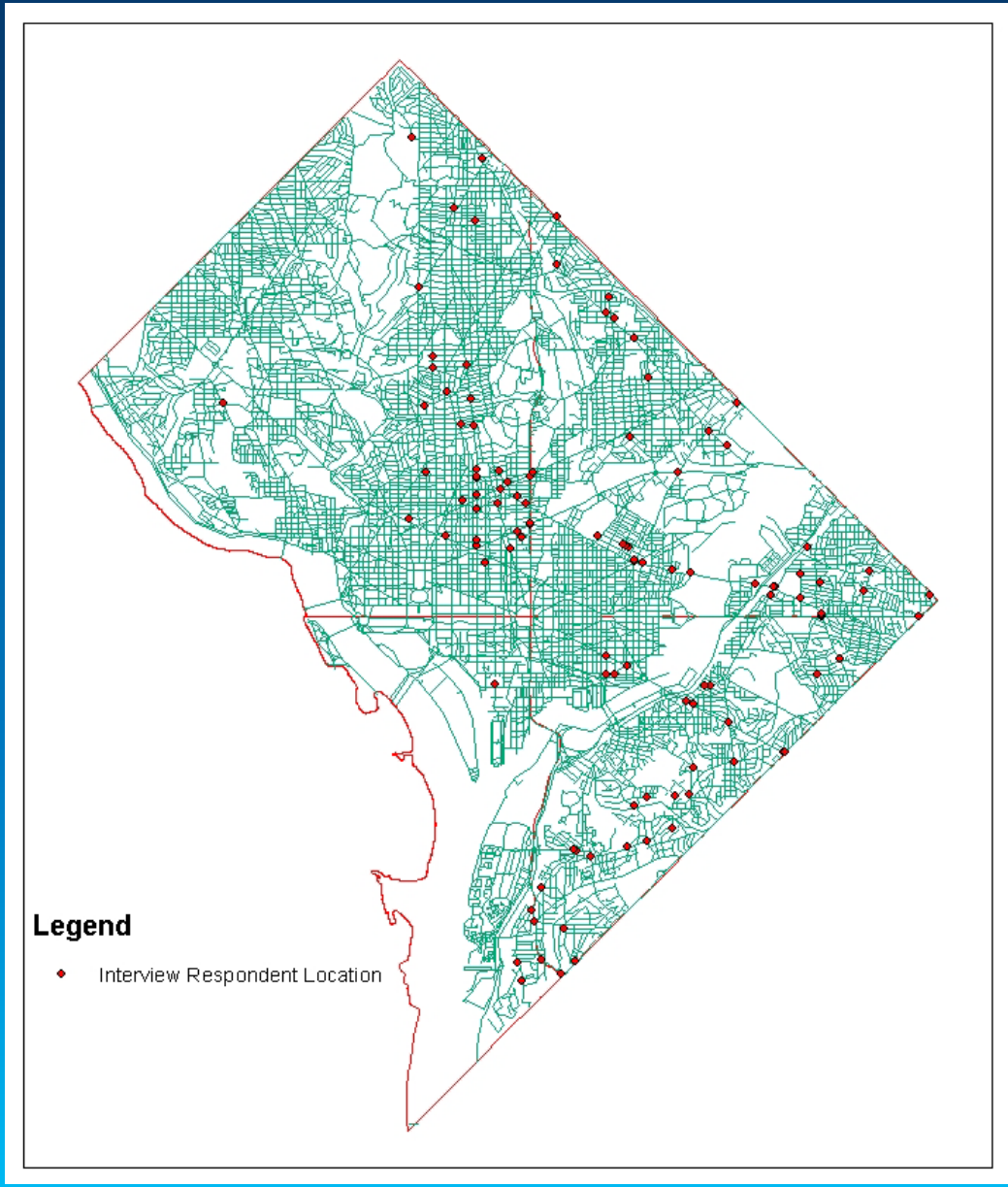
Methods – Primary Data Collection

Characteristic	Value	Frequency
Educational Attainment	None of the above (less than high school)	10 (6.6%)
	GED	15 (9.9%)
	High School Diploma	42 (27.6%)
	Some college	32 (21.1%)
	Associate's Degree	13 (8.6%)
	Bachelor's Degree	14 (9.2%)
	Master's Degree	4 (2.6%)
	Doctorate	1 (.7%)
Marital Status	Ever married	73 (48.0%)
	Never married	79 (52.0%)
Insurance Status	Yes	128 (83.7%)
	No	10 (6.5%)

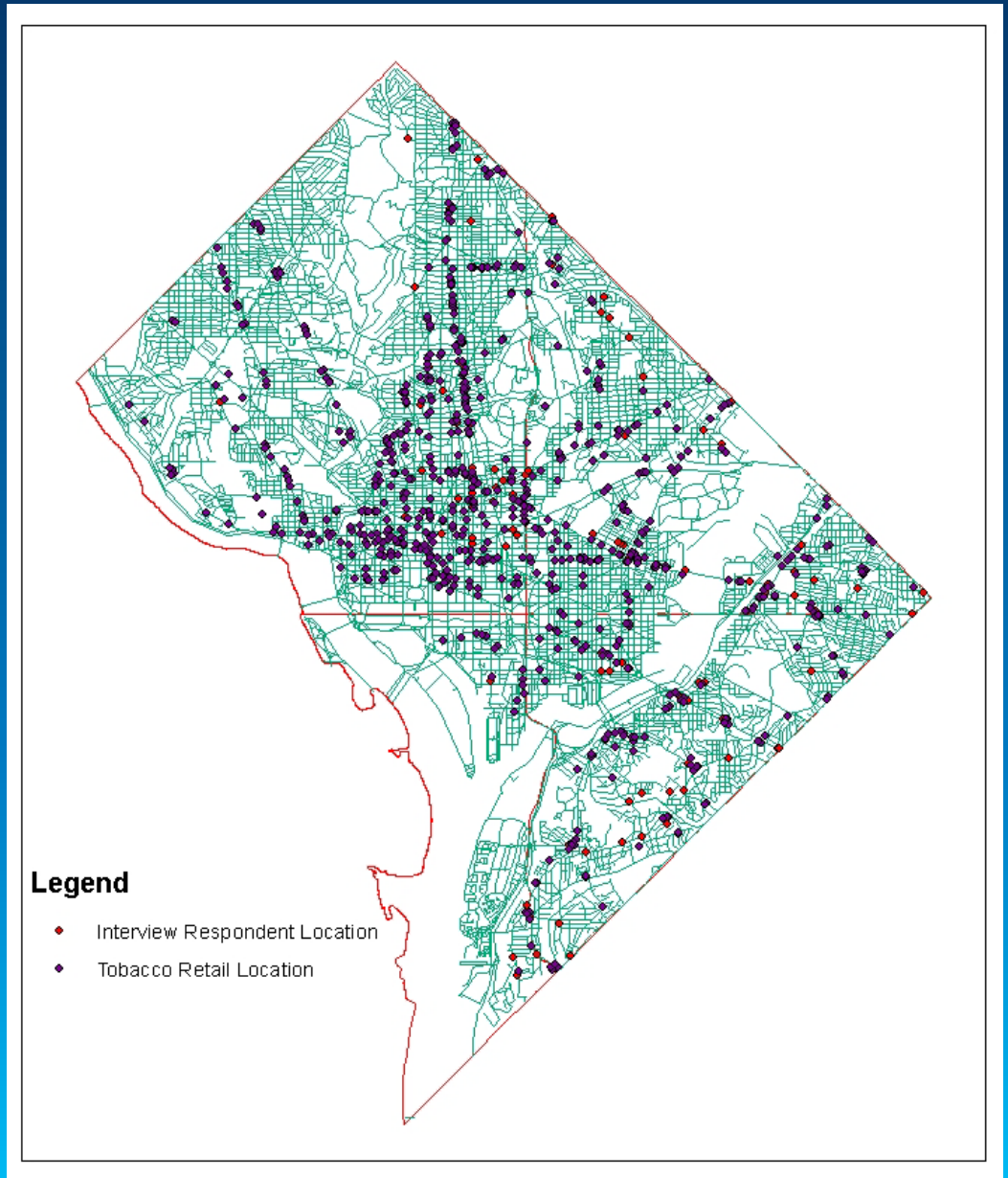
Methods – Secondary Data Sources

Original Data Source	Source for this project	Data Category	Website
Legacy Foundation Schroeder Institute	Legacy Foundation Schroeder Institute (P.I. Thomas R. Kirchner)	Tobacco Point-of-Sale data	Not applicable, proprietary database
U.S. Census	Neighborhood Info DC	Demographics	http://www.neighborhoodinfodc.org/
Behavioral Risk Factor Surveillance System (BRFSS)	Rand Corporation Analysis Report (Lurie, et al. 2008)	Ward Health Data	http://www.rand.org/content/dam/rand/pubs/working_papers/2008/RAND_WR534.pdf
Washington, DC Office of the City Administrator	Washington, DC Office of the City Administrator	Maps and political boundaries	http://data.dc.gov/Main_DataCatalog.aspx

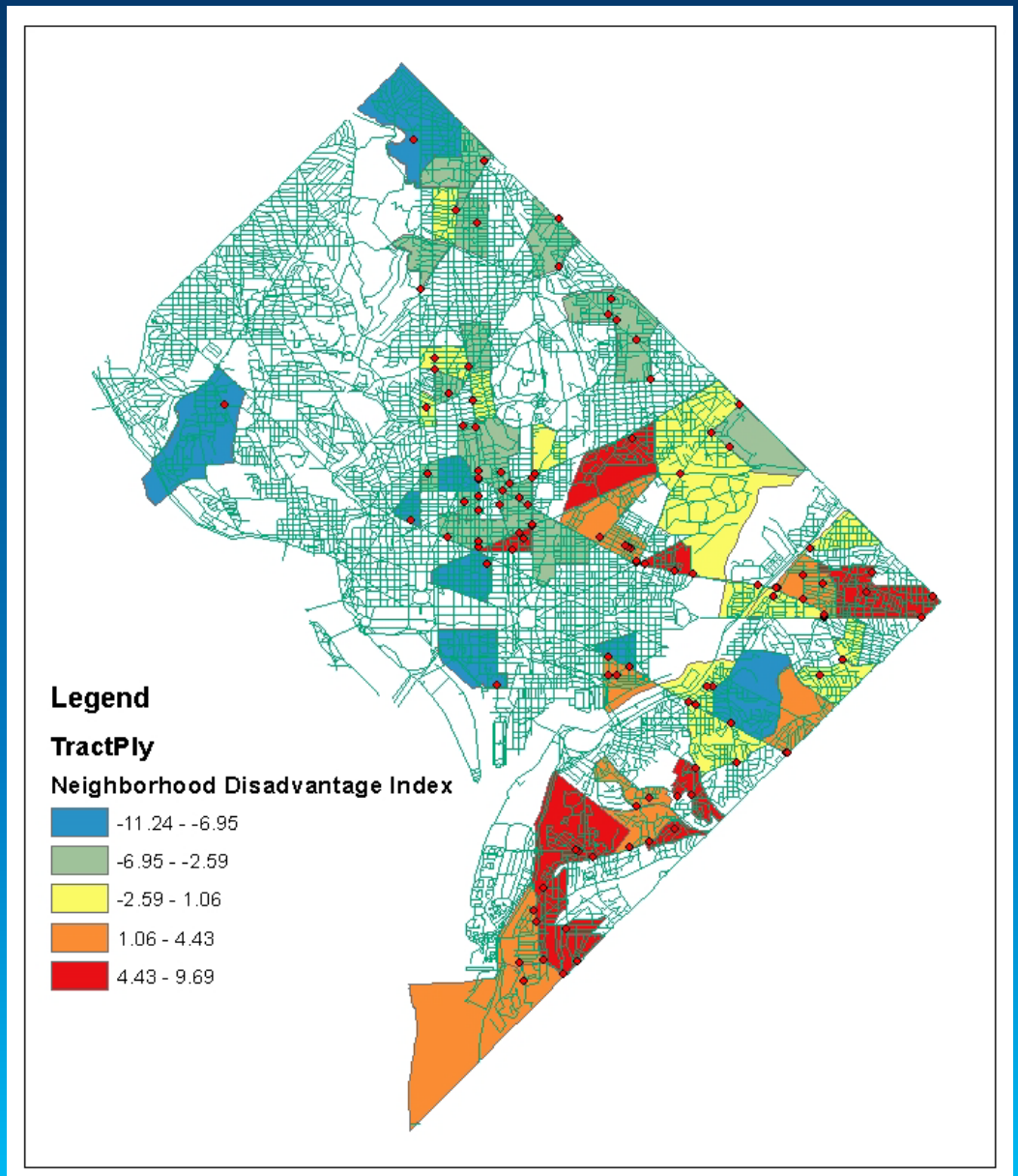
Results



Results


Legend
Interview Respondent Location
Tobacco Retail Location

Results


Legend
TractPly
Neighborhood Disadvantage Index
-11.24 - -6.95
-6.95 - -2.59
-2.59 - 1.06
1.06 - 4.43
4.43 - 9.69

Results – HLM Analysis

Fixed Effect	Coefficient	P-Value
Smoking Behaviors	-0.136	0.195#
Intent to Quit Tobacco	0.093*	0.012#
Socioeconomic Status	0.085*	0.049#
Medical Trust	0.090*	0.010#
Smoking-Related Diagnosis	-0.068	0.501#
Self-Reported Health Status	-0.049	0.781#

Note: The p-values above marked with a "#" should regarded as an approximation.
* p < .05, ** p < .01 (two-tailed test).

Results – HLM Analysis

Fixed Effect	Coefficient	P-Value
Pack Years	0.023*	0.014#
Usual Source of Care	0.055	0.854#
Quit Attempts	-0.014	0.183#
Gender (Female)	0.582*	0.039#
Age	0.003	0.814#
Marital Status	-0.669*	0.035#

Note: The p-values above marked with a "#" should regarded as an approximation.
* p < .05, ** p < .01 (two-tailed test).

Results – HLM Analysis

Fixed Effect	Coefficient	P-Value
For INTRCPT1		
Overall Social Support	-0.025*	0.022
Social Exposure to Smoking	0.023	0.088
Tobacco Retail Density	0.030	0.171
Neighborhood Disadvantage	-0.014	0.641
Neighborhood Health Status	0.002	0.881

* p < .05, ** p < .01 (two-tailed test)

Discussion

- **Individual factors were statistically significant predictors of tobacco cessation service utilization across several test models**
  - **Respondent's intentions to quit, socioeconomic status, medical trust, pack years, gender, and marital status were statistically significant**
  - **Smoking behavior and quit attempts in the past year were slightly non-significant**
- **Contextual Factors varied in statistical significance based on model**
  - **Social Support and Social Exposure were significant predictors of tobacco cessation service utilization**
  - **Tobacco Retail Proximity was statistically significant in reduced models**
  - **Neighborhood characteristics were non-significant**

Discussion

- **Study Limitations:**
  - **Small sample size**
  - **Homogenous in terms of race/ethnicity and insurance status**
  - **Limited specificity to neighborhood indicators and assessments in secondary data**

Lessons Learned

- **Collecting geographic information based on nearest intersection allowed estimating the neighborhood where the respondent lived without collecting street address.**
- **This helped ensure anonymity in the responses.**

Lessons Learned

- If there were multiple respondents who lived in a Census tract, the Join function sometimes did not copy those data into the final data set.
- Google Maps provided a good resource for longitude and latitude data of reported intersections.
- Using preprocessed datasets instead of raw Census or BRFSS files simplified data set construction and reduced the time of the study since there was no waiting for confidential data fields.

Future Research

- Conduct interviews with people in a social network and map their locations along with the initial respondent to determine clusters of characteristics and outcomes
- Collect more data about the person's daily activities and the built environment and neighborhood influences he/she encounters
- Add more qualitative data to the measures of the built environment

Acknowledgements

- DC Primary Care PBRN
  - Babafemi Adenuga, MD (President)
  - Finie Hunter-Richardson, M.P.H. (Research Coordinator)
  - HUH Doctors (Drs Howard Wilson, Robert Williams, Krishnan Narasimhan, and Oritsetsemaye Otubu) and medical staff (Nina Dickerson and Karla Espinoza)
  - HUH patients who participated
- Thomas R. Kirchner, Ph.D. and the Schroeder Institute/Legacy Foundation
- Forough Saadatmand, Ph.D. and the MIDARP staff (Research reported in this presentation was supported by National Institute on Drug Abuse of the National Institutes of Health under Award Number R24DA021470.)
- Robin Pugh-Yi