

JPSM/MPSDS Seminar Series, October 5, 2022

Should surveys produce more contextual features? Comparing contextual features by alternative definitions of neighborhoods

Shiyu Zhang ¹, Jessica Finlay ^{1,2}, Mao Li ¹, Robert Melendez ¹, Suzanne Judd ³, Michael Esposito ⁴, Philippa Clarke ^{1,2}

¹ Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI, USA

² Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI, USA

³ School of Public Health, University of Alabama at Birmingham, 1665 University Blvd, Birmingham, AL, USA

⁴ Department of Sociology, Washington University in St. Louis, St. Louis, MO, USA



INSTITUTE FOR SOCIAL RESEARCH
PROGRAM IN SURVEY AND DATA SCIENCE
UNIVERSITY OF MICHIGAN

Contextual features

Social and built environment features of neighborhoods around residential addresses

- Walkability
- Green space
- Accessibility of establishments and facilities
- Socioeconomic
- Demographic structure
- Crime and safety
- etc...

Contextual features

Social science and public health



Survey methodology

- Auxiliary variables on the entire sample

Contextual features

How is contextual features typically operationalized?

- Commonly defined for US Census geographies, like Census tracts
- Lowest level of geographic information available
- Convenient
- But, not defined for neighborhood research

→ There is a need to redefine neighborhoods so that the contextual features are specific and relevant to the studied individuals, *at least in theory....*

Contextual features

Empirically?

Census-based vs. case-specific contextual features: How much difference?

- Practical question
- Small difference → Use Census features for its convenience
- Big difference → Justify the production of case-specific contextual features

Research questions

1. How different are social and built environment features based on alternative definitions of neighborhoods?
 - Census tracts
 - Buffer
 - Respondent-informed neighborhood
2. How do these spatial differences affect associations between the contextual features and health-related outcomes?

Methods



INSTITUTE FOR SOCIAL RESEARCH

PROGRAM IN SURVEY AND DATA SCIENCE

UNIVERSITY OF MICHIGAN

Data

REasons for Geographic and Racial Differences in Stroke (REGARDS)

- nonprobability survey; sample from a commercial list
- White and black respondents
- 2003-2007: 30,000+ adults aged 45+ were recruited by telephone interviews
- Current analysis: 15,685 in second survey assessment between 2013 and 2016—the Second Survey to Assess Risk Factor (SSuRF).
- SSuRF includes 1) a computer-assisted telephone interview (CATI), 2) an in-home physical exam, and 3) a self-administered paper-and-pencil interview (PAPI)

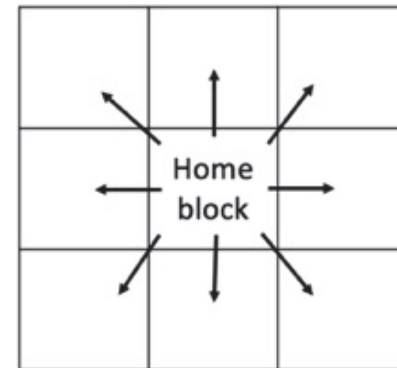
Three definitions of neighborhoods

For each SSuRF respondent

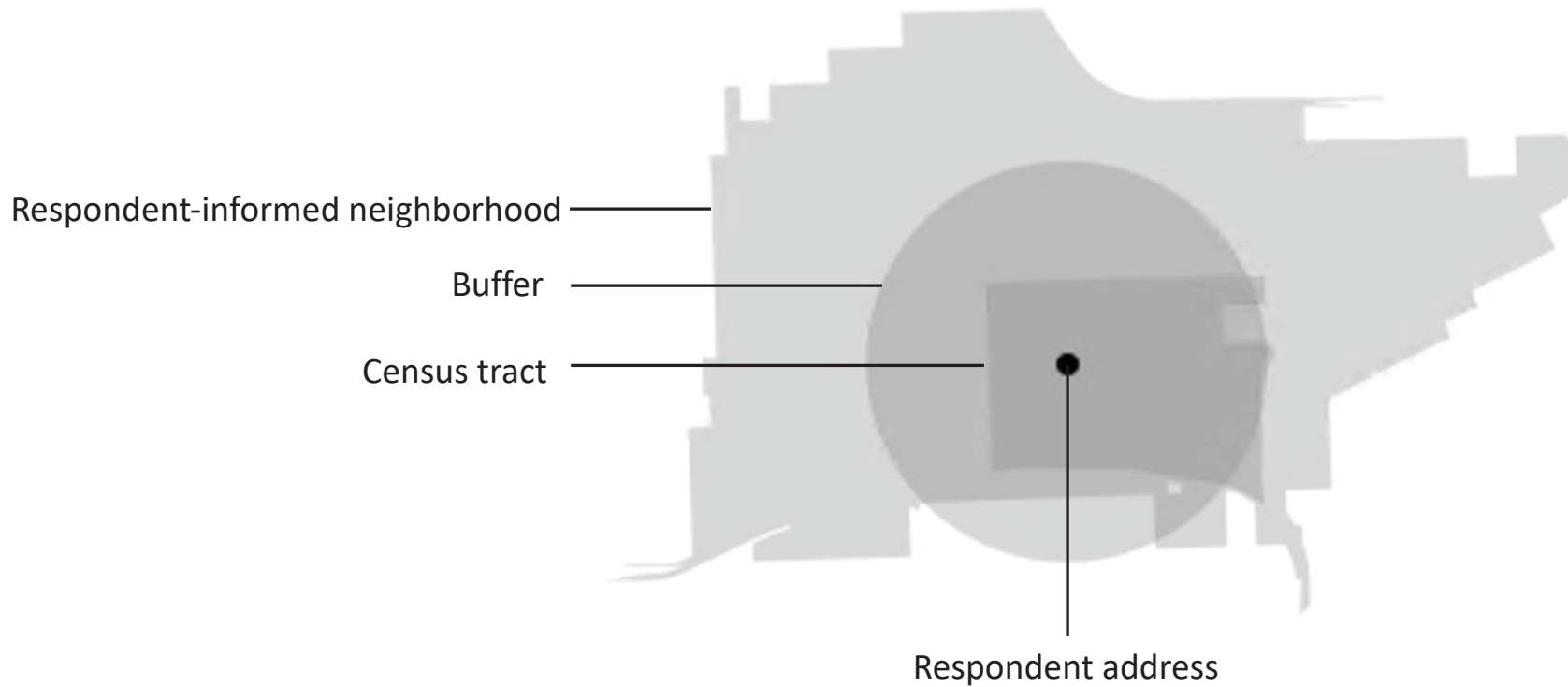
- 1. Census tract**
- 2. Buffer**
- 3. Respondent-informed neighborhood**

Based on self-reported neighborhood sizes & Census block geographies

“How many blocks are in the area that you think of as your neighborhood? By neighborhood, we mean the area around where you live and around your house. It may include places you shop, religious or public institutions, or a local business district. It is the general area around your house or apartment where you might perform routine tasks, such as shopping, going to the park, or visiting with neighbors”



Three definitions of neighborhoods



Contextual features

Sociodemographic structure

1. Neighborhood disadvantage
2. Affluence
3. Ethnic/immigrant concentration

Accessibility of active aging infrastructure

4. Parks
5. Recreational/sports places

Accessibility of intellectually stimulating places

6. Higher-education institutions
7. Libraries
8. Cultural institutions

Accessibility of social infrastructure

9. Civic and social organizations (e.g. veterans' membership organization)
10. Food and drinking places
11. Elderly services

Contextual features

Sociodemographic structure

1. Neighborhood disadvantage
2. Affluence
3. Ethnic/immigrant concentration

Data source:

Census tract level aggregated data based on the American Community Survey (ACS)

E.g.

Neighborhood disadvantage
= average rate of poverty, unemployment, female-headed families, and households receiving public assistance income

On Census tract:

ACS estimates

On buffer:

Overlaid ACS counts from tracts to buffers

On respondent-informed neighborhood:

Overlaid ACS counts from tracts to perceived neighborhoods

Contextual features

Accessibility of active aging infrastructure

4. Parks

Data source:

ParkServe database:
geographic data of urban parks across the US

On **Census tract:**

Presence of park

On **buffer:**

Presence of park

On **respondent-informed neighborhood:**

Presence of park

Contextual features

Accessibility of active aging infrastructure

5. Recreation/sports places

Accessibility of intellectually stimulating plac

6. Higher-education institutions

7. Libraries

8. Cultural institutions

Accessibility of social infrastructure

9. Civic and social organizations (e.g. veterans' membership organization)

10. Food and drinking places

11. Elderly services

Data source:

National Establishment Time-Series (NETS)
longitudinal database

On Census tract:

Counts of establishments

On buffer:

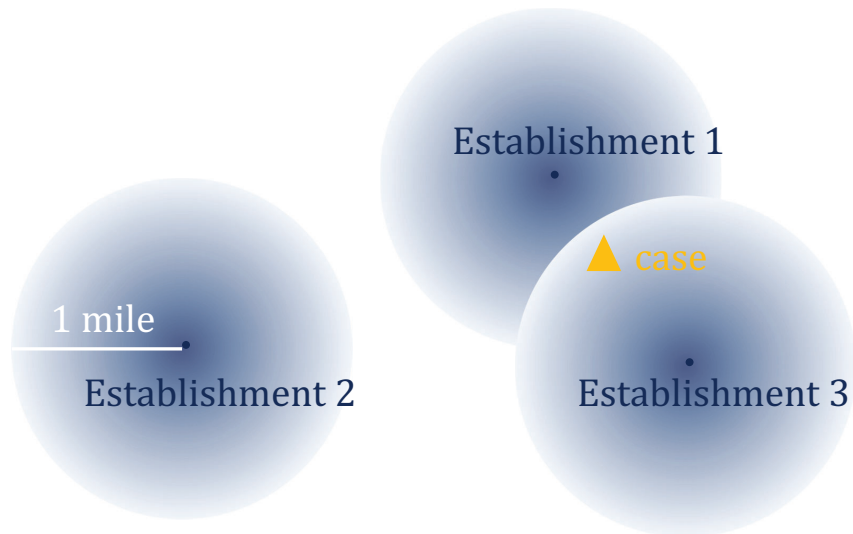
Kernel density estimation
(→ next slide)

On respondent-informed neighborhood:

Counts of establishment

Kernel density estimation

What is the case's accessibility of the establishments?



▲ locates within the influence radius of establishment 1 and 3.

Thus, for ▲, the accessibility level is the summed kernel density value of establishment 1 and 3

Outcome variables

Physical health

- Overweight (1/0)
- Diabetes (1/0)
- Hypertension (1/0)
- Stroke risk (continuous)
- Excellent health (1/0)
- Physical composite score (continuous)

Mental health

- Stress (continuous)
- Mental composite score (continuous)

Cognitive ability (continuous)

Activity level

- No exercise (1/0)
- # of walking days in past 7 days (continuous)
- Low physical activity (1/0)
- How often visit neighborhood (continuous)

Results



INSTITUTE FOR SOCIAL RESEARCH

PROGRAM IN SURVEY AND DATA SCIENCE

UNIVERSITY OF MICHIGAN

Research question 1

How different are the social and built environment features based on **alternative definitions of neighborhoods?**

Correlations between alternatively-defined features

Sociodemographic indices

		Buffer	Informed
Disadvantage	Tract	0.91	0.88 <i>weak</i>
	Buffer		0.94 <i>strong</i>
Affluence	Tract	0.93	0.92 <i>weak</i>
	Buffer		0.94 <i>strong</i>
Ethnic/ immigrant	Tract	0.94	0.92 <i>weak</i>
	Buffer		0.94 <i>strong</i>

Accessibility of...

		Buffer	Informed
Active aging – sports places	Tract	0.37	0.34 <i>weak</i>
	Buffer		0.41 <i>strong</i>
Intellectually stimulating	Tract	0.29	0.27 <i>weak</i>
	Buffer		0.37 <i>strong</i>
Social infrastructure	Tract	0.29	0.24 <i>weak</i>
	Buffer		0.37 <i>strong</i>

Research question 2

How much do these spatial differences of neighborhoods affect **associations** between contextual features and individual health outcomes?

Research question 2

Outcome variables

Physical health

- Overweight (1/0)
- Diabetes (1/0)
- Hypertension (1/0)
- Stroke risk (continuous)
- Excellent health (1/0)
- Physical composite score (continuous)

Mental health

- Stress (continuous)
- MCS (continuous)

Cognitive ability (continuous)

Activity level

- No exercise (1/0)
- # of walking days in past 7 days (continuous)
- Low physical activity (1/0)
- How often visit neighborhood (continuous)

Outcome ~ contextual feature (tract/ buffer/ informed)

+ demographic variables (gender, age, race, education, marital status)

- A total of 429 models (= 11 features * 3 neighborhood def. * 13 outcomes)
- Whether the contextual predictors are significant?
Whether the significance levels are the same across neighborhood definitions?

Research question 2

- A total of 429 models (= 11 features * 3 neighborhood def. * 13 outcomes)

	Contextual feature 1	feature 2 ...
Outcome 1	Sig based on tract	...
	Sig based on buffer	...
	Sig based on informed NB	...
Outcome 2	Sig based on tract	...
	Sig based on buffer	...
	Sig based on informed NB	...
...		



	Contextual feature 1	feature 2
Outcome 1	Same sig	
Outcome 2	Different sig	
...		

	Sociodemographic			Active aging		Intellectually stimulating			Social infrastructure		
	1 Disadv.	2 Afflu.	3 Ethnic	4 Park	5 Sports	6 Educ.	7 Library	8 Culture	9 Civ/Soc	10 F/Drink	11 Service
BMI	***	***		**	*	***	***	***	**	**	*
diabetic	**	***		***	***		*	**	**		*
hypertension	***	***	*		***	*	*	*	***	**	
Stroke risk	*	***		***	**		**	**	*	**	
Excellent health	***	***	*		***	**	**	*	**	**	
PCS	***	***		**	**	*	*	*	*		
Stress	***	***	*		*		**	*	**	*	*
MCS	**	*			*	***	**	**	*	**	
Cognitive ability	***	***	***	***	*	***	***	***	***	***	*
No exercise	**	***			*			*			
Walk days	***	***		*	***	***	***	***	***	***	**
Low activity	***	***		***	**	*	**	**	**	**	*
Nb. visit	***	***	**	*	***	*	**	**	***	**	**

	Sociodemographic			Active aging		Intellectually stimulating			Social infrastructure		
	1 Disadv.	2 Afflu.	3 Ethnic	4 Park	5 Sports	6 Educ.	7 Library	8 Culture	9 Civ/Soc	10 F/Drink	11 Service
BMI		***		**	***	***	***	***	**	***	***
diabetic	**	***		***	***	***	*	**	**	***	***
hypertension	***	***	***	***	***	***		***	***	***	***
Stroke risk	*	***		***	***	***		***	***	***	***
Excellent health	***	***	*	***	***	***		***	***	***	***
PCS	***	***		***	***	***		***	***	***	***
Stress	***	***	*	***	***	***		***	***	***	***
MCS	***	*		***	***	***		***	***	***	***
Cognitive ability	***	***	***	***	***	***		***	***	***	***
No exercise	**	***		***	***	***		***	***	***	***
Walk days	**	***		***	***	***		***	***	***	***
Low activity	***	***		***	**	*	**	**	**	**	*
Nb. visit	***	***	**	*	***	*	**	***	***	***	**

Significance count:

- Tract: 55
- Buffer: 84
- Informed neighborhood: 114

Significance agreement:

- Tract= Buffer: 97
- Buffer= Informed: 91
- Tract= Informed: 78

Census tract

- Least relevant
- Fewest significant results

Buffer

- “Middle-ground”

Respondent-informed
neighborhood

- Most relevant
- Most significant results

Agree least
Weakest correlation

Conclusions and other thoughts

- Definition matters
- Substantive research: respondent-informed neighborhood features
- Survey research: buffer features
- The influence of data structure and survey mode
- *Units* of measurements. What about *methods* of measurement?
- Take-away message:
Empirical differences between tract features and case-specific features justify additional investment in data production? We think so.

Additional slides

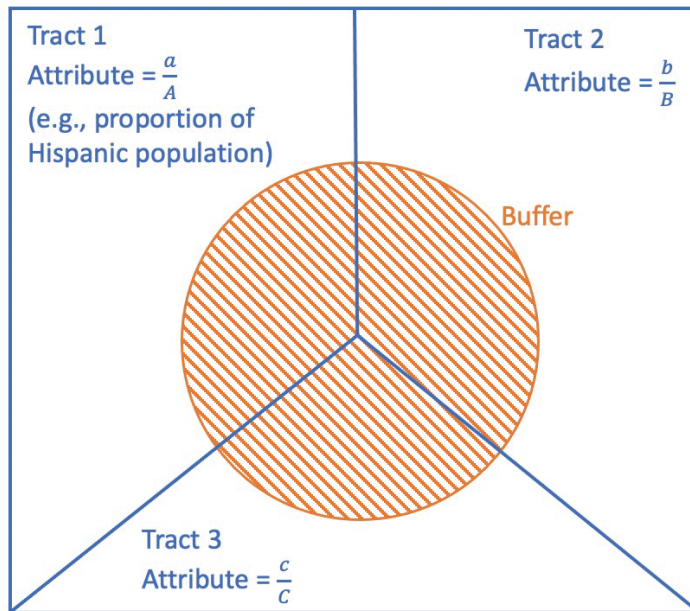


INSTITUTE FOR SOCIAL RESEARCH

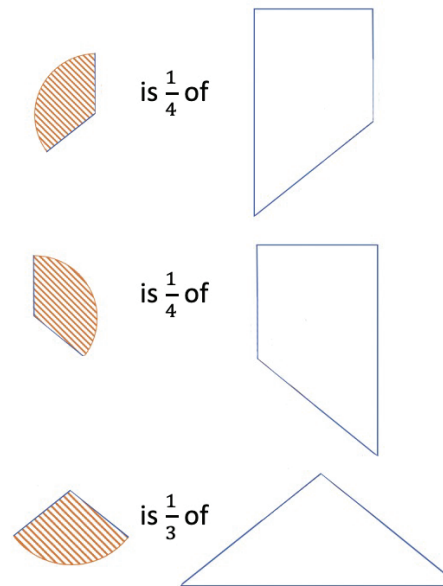
PROGRAM IN SURVEY AND DATA SCIENCE

UNIVERSITY OF MICHIGAN

Overlay



Suppose that



then

$$\text{measure of buffer} = \frac{\frac{1}{4}a + \frac{1}{4}b + \frac{1}{3}c}{\frac{1}{4}A + \frac{1}{4}B + \frac{1}{3}C}$$