Using Spatial Aggregation Method outcome to explain the influences of built environment on health profile

Yuanyi Shen[1], Mengqi Zhong[2], Qingla Zhang[3], Jiatian Bu[4], Yifan Yu[5]

[1]*Tongji University, j03u624@tongji.edu.cn*
[2]*Tongji University, 476333440@qq.com*

Abstract: Health profile is becoming more significant in Developing country such as China.

Research into the effects of the built environment on health has increased dramatically in recent years.Researchers are trying to understand and research explores the role of the built environment and transport system on an individual's health.Data from recent reviews have shown associations between environmental features, such as presence of sidewalks, proximity to parks and presence of certain types of food outlets, and outcomes such as physical activity. However, results from many studies conducted have been non-significant. A primary cause of these non-significant is due to how neighborhood areas are defined, which directly affects how the built environment variables are calculated in geographic information systems.

In this paper tests to what extent the potential impacts on regression analysis resulting from different data aggregation methods are well documented in spatial studies by varying the initial geographical scale of analysis which is primarily referred to as the modifiable aerial unit problem.

As explained earlier, the focus is on reducing the error caused by the modifiable aerial unit problem by introducing a data aggregation method.

Individual health and lifestyle data are obtained from the survey of income, and labor dynamics in census figures of China, and the relationship between the built environment and health profile is evaluated by using a discrete choice model.

The intended results is identify which variable is more closely related to health status by the proposed aggregation method is evaluated across three spatial scales

Keywords: built environment, spatial, geographic, health profile, regression analysis

Introduction

Health outcomes have been a central concern in evaluations of quality of life (Brazier, Ratcliffe, Saloman, and Tsuchiya 2017). Obesity prevalence has increased at a unhealthy rate over the last three decades in industrialized countries. This sudden increase in unhealthy is more likely to be related to environmental changes than to biological changes. Several empirical studies document a relationship between built form and obesity in adults (Frank et al., 2004;Lopez, 2004; Lopez-Zetina et al., 2006; Xu et al., 2005). Green infrastructure has emerged as a topic of significant interest in urban and regional planning. A lot of scientists and public health experts determine that environmental change could help health—for example, by restricting the number of fast-food restaurants in a neighborhood (Minkler, Wallerstein, andWilson 2008). Different population and urban form factors also have an impact on walking and weight. Other studies have shown that store density is related to physical activity (Atkinson et al., 2005, Besser and Dannenberg, 2005, Handy et al., 2005, Frank et al., 2005). Than researchers are trying to understand the environmental variables that affect health. Land use change and urbanization have changed the architectural pattern of metropolitan areas. Studies have found land use, plot size and health-related variables (Amir Samiz, Abolfazl Mohammadian, Seyedali Madanizadeh 2009). One related growing stream of research explores the role of the built environment on physical activity (Samimi, A., A. Mohammadian, and S. Madanizadeh 2009) and the individual's health (Coffee T. N 2005). However, results from many of the studies conducted have been contradictory (Zhang, M., and N. Kukadia 2005). One of the primary causes of these contra- dictions can be attributed to the way the neighborhoods are defined, which directly affects how the built environment variables are calculated in geographical information systems (GIS). Spatial analysis is the foundation of many studies within the transportation field. Spatial analysis implicitly defines some of the underlying assumptions of a study. A typical spatial analysis is the aggregation of data within geographic boundaries. The distribution of data can be clearly seen through aggregation.

Materials and methods

Study population

This section describes the study area, data source, and processing procedures used in this study. The study area is defined by the Shanghai Statistical Division, which is a large administrative zone that encompasses the city of Shanghai, China. The study area includes 2,253,525 participants who responded to the survey (referred to here as "respondents") within 5432 census collection districts (CCDs).

The health used in this study came from the Shanghai Statistical Division (SSD) survey, which began in 2010 and surveyed the Shanghai population. The data were subsequently processed in ArcGIS and spss to calculate a number of environment variables (introduced in the next section) that represented the urban characteristics of Shanghai, China

Health outcome assessment

Survey respondents were asked: "Would you say that in general your health is excellent, good, fair or poor?" SSD used this measure of self-rated health -- used extensively in the literature, and shown to be a good marker of combined physical and mental health (Idler and Angel, 1990, Idler and Benyamini 1997) -- as a continuous measure, coded such that increasing values denote better self-rated health.

Built Environment Variables

The"six Ds"principle, which accounts for the main ways in which the built environment is expected to influence travel behavior, was used as a guideline for the selection and development of the built environment variables included in this study. The six Ds principle is composed of density, diversity, design, destination, distance to transit, and demand management (Ewing, R., and R. Cervero 1998).

The destination principle was considered by using the variable "quantity and distance to supermarket" A code was developed in Python to obtain the walking distance and quantity to the grocery stores and sports facilities closest from Gaode Maps. These are the two major points of interest affecting health. In order to convert these two variables to the CCD zones, overlapping two layers and CCD layer for range segmentation to calculate the number of shops and sports facilities in each CCD range. Greenland data from Shanghai Surveying and

Mapping Department. The green space is divided according to each CCD range, and the green area of each CCD range is calculated.

.Method

The purpose of this study is to evaluate the association of urban environment characteristics and health in Shanghai, China, and to examine the importance of spatial aggregation methods in the significance of this association. The residential location of individuals is preaggregated to CCDs in the data used in this study. Although CCDs are relatively small zones (resident committee management), it is not possible to de ne the neighborhood area for each individual by using a road buffer from the address of each individual. Thus, the only link between the data and the built environment variables is the CCD. The CCDs were defined for the purposes of data collection during the census.

The ideal aggregation methodology would de ne the neighborhood area on the basis of how each individual interacts with the physical environment surrounding his or her place of residence; this method would remove the associated MAUP errors. This problem is referred to as the uncertain geographic context problem, which requires that the travel trajectory of each individual represent the urban area of influence within which an individual would have interactions (Kwan, M.P 2012). Using the self-assessment health data of the sixth census for spatial clustering, we can know the relationship between health status and built environment through clustering results. Many studies have shown that the more points of interest(POI) and green spaces and health facilities near the place of residence have a stronger link to health outcomes(Crawford， Thomas W 2014). The following steps are used to calculate the built environment variables using the suggested aggregation method:

1.Overlapping CCDs and land-use layers containing green spaces;

2. Remove the protective green space next to the road;

3. The green map layer is overlapped by the CCD range line for regional division;

4. Calculate the green area near each neighborhood committee;

5. Overlay the layers aggregated with the self-assessment health survey;

6. Calculate whether there is any correlation with self-assessment health data;

Statistically, correlations between multiple variables (statistics) are detected by correlation analysis. For example, rice yield is usually related to soil fertility. If the statistic of the analysis is the same attribute variable of different observation objects, it is called "autocorrelation". Therefore, the so-called spatial autocorrelation is to analyze these spaces by statistically studying the spatial degree of autocorrelation between a spatial unit and its surrounding units in space. Characteristics of unit space distribution phenomena. Therefore, the local spatial autocorrelation of health data is first self-assessed to understand the spatial distribution of health. The regional Moran’s I measures the degree to which each i is related to each j within the range of distance d. The formula is as follows:

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_j w_{ij}(x_j - \bar{x})$$

Xi, Xj are observations of spatial positions i and j, and w ij represents the neighborhood of spatial positions i and j.

Statistical analyses

Within each areal neighborhood definition, we calculate Pearson correlations among the four environmental variation. This relation is because the defined neighborhood area is more likely to reflect the area in which an individual had direct interactions as compared with an administrative boundary or a straight-line buffer (Duncan, M. J., E. Winkler, T. Sugiyama, E. Cerin, L. duToit, E. Leslie, and N. Owen.). To measure the association between the built environment and health, a logit model is developed. To measure the association between the built environment and obesity, a one-dimensional logit model was developed. Four sets of environmental variables were modeled and saliency calculated, taking into account the potential relationship between individual variables and health. Healthy input - output model.

Results

Characteristics of survey population

A majority of participants reported at least good self-rated health: 151019 reported 'excellent' self-rated health, 150675 reported 'good,' 31248 'fair,'12930 'poor,'and 1907933 unfilled.

Associations with self-reported health, environment variable

In each dataset, Pearson correlations were lower for green area (Table 1). The number and health status of sports facilities are significant, with the highest correlation between the two variables. Although the number of health and stores is significant, the two variables are negatively correlated. Table 1 shows the results across the four models as discussed previously. From the green area variables not has a association with health, which may seem counterintuitive at the first glance (since green area is considered healthy). From the built environment variables, the number of Sports Facilities has a positive association with health. In general, the area of green space has no significant impact on health, but the number and distance of sports facilities have an impact on health.

Table 1. Pearson correlations of health by environment variables.

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.253	.072		-3.534	.000
	Green Area	-5.550E-7	.000	-.021	-1.561	.119
	Restaurant	-.004	.001	-.071	-3.188	.001
	Sports Facilities	.022	.005	.090	4.123	.000
	Population"	.000	.000	.213	15.320	.000
	Subway Station	-.149	.141	0.014	-1.052	.293
a. Dependent Variable: Health						

Cluster analysis of self-rated health

This study proposed, applied, and tested an aggregation methodology for anonymous surveys in which the self-rated health of individuals is preaggregated to an existing administrative boundary. 151019 of participants reported at excellent self-rated health that use localized spatial autocorrelation analysis(Figure 1).The low-low clusters are mainly concentrated in the central area of Inner Ring Line of Shanghai. Residents living in the city centre are particularly bad at their health, and than it can be seen that green space has no effect on health.

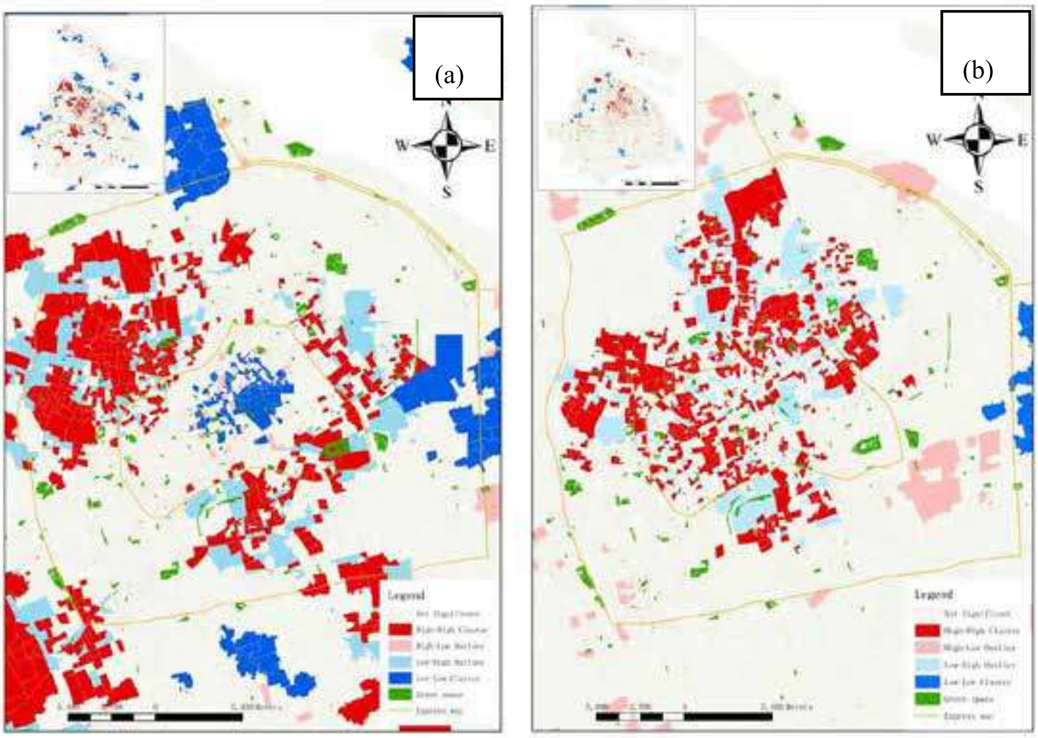

FIGURE 1 Aggregation clustering:(a) 'excellent'self-rated health within CCD boundary; (b) 'poor'self-rated health within CCD boundary

Discussion of results

The study assessed the association between built environment and health in Shanghai, China. A significant data collection and processing effort was conducted to obtain GIS data necessary for the evaluation of the built environment variables. The variables were selected on the basis of the six Ds principle, which considers how the urban environment may affect travel behavior. Results from the study indicate that there is an association between built environment and health, particularly when the number of Sports Facilities. But residents living in the city centre are less satisfied with their health than those living outside the city. There may be three reasons,(1)More demanding for your health; (2)The way to go out is mostly negative travel; (3) Getting food is more convenient. Through the aggregation clustering of health self-assessment data, the results show the distribution of residents' health status in Shanghai.

Conclusion

Our study aimed to understand the importance of: (1) know what environmental variables affect the health of residents in the city; (2) aggregation of vegetation data into the same areal units when seeking to evaluate potential effects of neighborhood greenness exposure on health. Using individual-level data obtained from a survey of Shanghai residents, we compared associations between self-rated health and greenness exposure.

The size and shape of neighborhoods as perceived by each individual (i.e., self-defined) could differ due to demographic characteristics of different subgroups within a population.Anonymous surveys tend to aggregate the residential location of survey respondents to predefined zones. These zones are typically not designed to capture the urban characteristics of a neighborhood. The issue associated with defining the spatial aggregation boundary is referred to as the MAUP and has been shown to have dire consequences, particularly in regression analysis. In this study a variety of data sets describing the urban characteristics of Shanghai, China, are collated, processed, and analyzed. And contrary to expectations, we found not associations between green space and self-reported health, And the number of sports facilities significantly affects health. A previous US Study found that areas around home contained more fitness facilities than non-home areas(Hurvitz PM, Moudon AV.2012). It can be considered that the health status of residents outside the city center is better, because the residential area is mostly outside the city center, and the sports facilities are located near the residential area, so the residents living in the center of the city have a higher health index.

The current study has some limitations. Due to the cross-sectional nature of our research, reverse causality is possible, and it is impossible to truly identify all the variables that affect the state of self-health, and only actually reflect the distribution of health status of the city.

References

Amini, N., T. Rashidi, L. Gardner and S. T. Waller (2016). "Spatial Aggregation Method for Anonymous Surveys: Case Study for Associations Between Urban Environment and Obesity." Transportation Research Record: Journal of the Transportation Research Board 2598(1): 27-36.

Brazier, J., Ratcliffe, J., Saloman, J., & Tsuchiya, A. (2017). "Measuring and valuing health benefits for economic evaluation". Oxford: Oxford University Press.

Coffee, N. T., N. Howard, C. Paquet, G. Hugo, and M. Daniel (2013). "Is Walk- ability Associated with a Lower Cardiometabolic Risk? " Health & Place, Vol. 21, 2013, pp. 163–169

Crawford, Thomas W.; Pitts, Stephanie B. Jilcott; McGuirt, Jared T.; et al (2014). "Conceptualizing and comparing neighborhood and activity space measures for food environment research . "HEALTH & PLACE Volume: 30 Pages: 215-225 Published: NOV 2014

Duncan, M. J., E. Winkler, T. Sugiyama, E. Cerin, L. duToit, E. Leslie, and N. Owen (2010). Relationships of Land "Use Mix with Walking for Trans- port: Do Land Uses and Geographical Scale Matter? " Journal of Urban Health: Bulletin of the New York Academy of Medicine, Vol. 87, No. 5, 2010, pp. 782–795.

Ewing, R., and R. Cervero (2010). " Travel and the Built Environment. " Journal of the American Planning Association, Vol. 76, No. 3, 2010, pp. 265–29

Eisenman, T. S. (2013). "Frederick Law Olmsted, Green Infrastructure, and the Evolving City." Journal of Planning History 12(4): 287-311.

Ewing, R., G. Meakins, S. Hamidi and A. C. Nelson (2014). "Relationship between urban sprawl and physical activity, obesity, and morbidity - update and refinement." Health Place 26: 118-126.

Frank, L. D. and P. O. Engelke (2001). "The built environment and human activity patterns: Exploring the impacts of urban form on public health." Journal of Planning Literature 16(2): 202-218.

Reid, C. E., L. D. Kubzansky, J. Y. Li, J. L. Shmool and J. E. Clougherty (2018). "It's not easy assessing greenness: A comparison of NDVI datasets and neighborhood types and their associations with self-rated health in New York City." Health & Place 54: 92-101.

Samimi, A., A. Mohammadian and S. Madanizadeh (2009). "Effects of transportation and built environment on general health and obesity." Transportation Research Part D: Transport and Environment 14(1): 67-71.

Frank, L. D., J. Kerr, J. F. Sallis, R. Miles and J. Chapman (2008). "A hierarchy of sociodemographic and environmental correlates of walking and obesity." Prev Med 47(2): 172-178.

Hurvitz PM, Moudon AV (2012). "Home versus nonhome neighborhood: quantifying differences in exposure to the built environment. " Am J Prev Med. 2012;42(4):411–7.

Kwan, M.-P. " How GIS Can Help Address the Uncertain Geographic Context Problem in Social Science Research. " Annals of GIS, Vol. 18, No. 4, 2012, pp. 245–255.

Leal, C., and B. Chaix. "The In uence of Geographic Life Environments on Cardiometabolic Risk Factors: A Systematic Review, a Methodological Assessment, and a Research Agenda. "Obesity Reviews, Vol. 12, No. 3, 2011, pp. 217–230.

Leal, C. and B. Chaix. "The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda." Obes Rev,2011, 12(3): 217-230.

Minkler, M., N. Wallerstein, and N. Wilson. (2008). "Improving Health through Community Organization and Community Building". In Health Behavior and Health Education: Theory, Research and Practice, 4th ed., edited by K. Glanz, B. Rimer, andK. Viswanath, pp. 279–311. San Francisco : Jossey-Bass.

Macdonald, L. (2019). "Associations between spatial access to physical activity facilities and frequency of physical activity; how do home and workplace neighbourhoods in West Central Scotland compare?" Int J Health Geogr 18(1): 2.

Reid, C. E., L. D. Kubzansky, J. Y. Li, J. L. Shmool and J. E. Clougherty (2018). "It's not easy assessing greenness: A comparison of NDVI datasets and neighborhood types and their associations with self-rated health in New York City." Health & Place 54: 92-10

Wang, F., M. Wen and Y. Xu (2013). "Population-Adjusted Street Connectivity, Urbanicity and Risk of Obesity in the U.S." Appl Geogr 41: 1-14.

Zhang, M., and N. Kukadia. (2005)Metrics of Urban Form and the Modi able Areal Unit Problem. "In Transportation Research Record: Journal of the Transportation Research Board ", No. 1902, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 71–79.