An exploratory spatial analysis to identify the appropriate scale and potential risk factors for understanding obesity distributions

YQ Tian[1], TS Troppy[2], W Dripps[1], and D Chalfoiux[3]

[1] Department of Environmental, Earth and Ocean Sciences, University of Massachusetts, Boston, MA 02125 Tel +01 617 2875285; Fax +01 617 2873929; Email yong.tian@umb.edu

[2] Boston Public Health Commission, Communicable Disease Control Division, 1010 Massachusetts Avenue (2nd Floor), Boston, MA 02118

[3] Boston Inspectional Services Department, 1010 Massachusetts Avenue (4th Floor), Boston, MA 02118

INTRODUCTION

Over the last two decades the United States has experienced an increase in obesity (as defined by a Body Mass Index (BMI) ≥25) that cuts across age, race, and gender (Anderson, 2000). An estimated 300,000 deaths per year in the US alone are currently associated, in some fashion, with being overweight, making the issue a significant concern for legislators, nutritionists, and policy makers. The total costs attributable to obesity were an estimated $117 billion (55% direct and 45% indirect costs) in 2000 alone, equal to approximately 8 percent of the US national health care budget (Wolf, 2001).

Most of the current research on obesity comes from the medical fields and, as such, focuses on medical treatment of obesity patients, including surgery and administering pharmaceuticals. A number of scientists have begun work identifying the various risk factors which may cause or worsen obesity in efforts to help reach the ideal solution of obesity prevention. Aside from possible genetic influences, different human activities including diet and physical exercise clearly affect obesity, but to make matters complicated, many of these activities are influenced not only by personal choice, but also by access, financial status, and other circumstantial constraints. Dietary choices, for instance, may be influenced by knowledge of the causative and preventative effects of certain foods, educational attainment (Shea, et al., 1991; Shimakawa et al., 1994), the cost of food, and the availability of different foods, among many other factors (Morland et al., 2002). Obesity prevention requires changes in individual behavior patterns as well as eliminating environmental barriers to healthy food choices and active lifestyles, both difficult to achieve in today's urban environments.

Social scientists have only recently begun to weigh in on the obesity issue, looking for causal spatial linkages between obesity and various social and physical parameters, many of which are collected for other uses. The challenge has been merging disparate datasets and data sources, identifying the appropriate spatial and temporal scales for analysis, determining the ideal sample size, and pinpointing the major spatial drivers linked to obesity. Obesity may be correlated to many commonly collected demographic and social variables like population density, economic status, ethnicity, age, and educational attainment, but it may also be linked to less traditionally collected variables like distance to a local recreation center, access to public transportation, and distribution and patterns of fast food restaurants and grocery stores. To better understand the obesity epidemic, researchers need to collect more data on obesity distributions, but perhaps as important identify what types of spatial data we should be collecting to interpret and understand these distributions.

In this paper, we present a very preliminary study that attempts to identify factors that might have causal relationships to obesity and evaluate the current data available to researchers on a neighborhood scale for possible improvement of data collection and compilation of research datasets intended to study obesity. The study makes use of data collected by the Boston Department of Public Health, which routinely collects obesity information as part of their city-wide health reports, as well as information from various local, state, and national agencies. The intent of this study is to link the obesity data to existing spatial physical and socio-economic databases to look for spatial correlations that may provide a better understanding as to the factors that influence obesity with the aim of helping prevent obesity rather than relying on medical treatments as most current efforts are now presently directed.

Our working hypothesis is that in addition to genetic factors, obesity may be linked to various social, economic, and physical drivers including less available datasets like the distribution, type, and density of restaurants or choice and access to physical exercise facilities for instance. Owing to the complexity of the relationship between human activities and decisions and public health phenomena, our efforts focus on a basic exploratory spatial analysis. We are using Geographic Information Systems (GIS) to test relationships of spatially distributed objects in the area of interest and how these objects can serve as a tool in identifying risk factors for obesity. Through the use of GIS we hope to provide the tools needed to store, analyze, and display obesity information about Boston and the neighborhoods that makeup the city. Obesity distributions do not appear to be random; as such, we strive to better understand this distribution by searching for the various spatial drivers that can explain and perhaps help predict this distribution (Rothman, 1990; Wu et al., 2004).

Methodology

Study site

The neighborhoods of Boston were selected as a study site based on the abundance of available data from the various local and state agencies. The City of Boston, with a population of 589,141, consists of 16 distinct neighborhoods (Allston/Brighton, Back Bay, Charlestown, East Boston, Fenway, Hyde Park, Jamaica Plain, Mattapan, North Dorchester, North End, Roslindale, Roxbury, South Boston, South Dorchester, South End, and West Roxbury) (Figure 1).

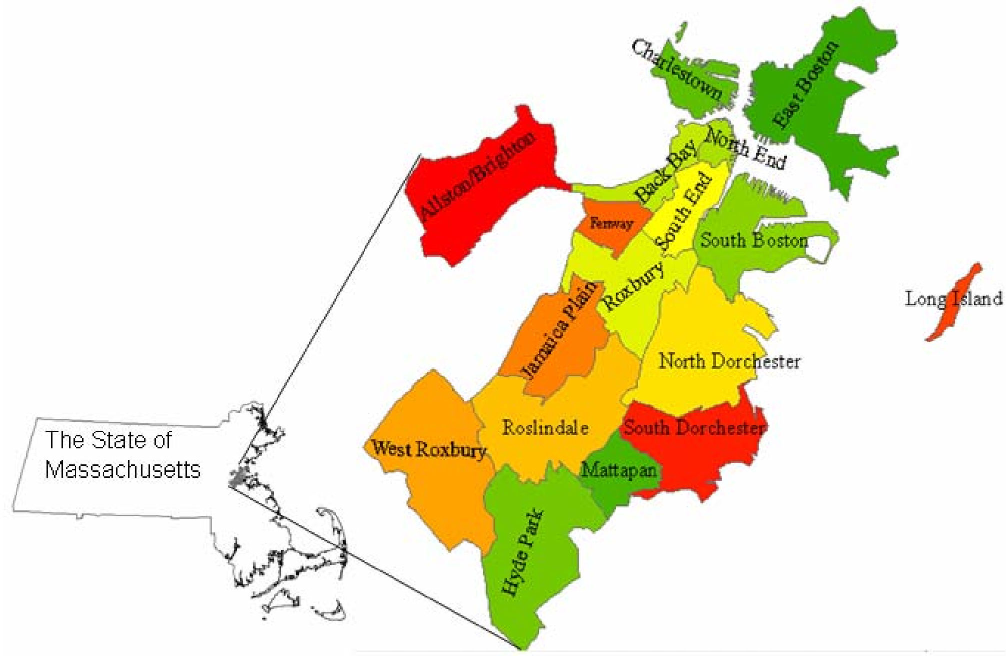

Figure 1. Boston neighborhoods selected for the study

Data description

Health data for the 16 neighborhoods were obtained from the National Centers for Disease Control and Prevention (CDC) Behavioral Risk Factor Surveillance System (BRFSS) Survey. The BRFSS is the world's largest telephone survey which tracks health risks in the United States. Information from the survey is used to improve the health of the American people. The survey includes incidence rates (%) of cancer and heart disease mortality and percentages of individuals who are obese and who have high blood pressure (Table 1). Each health indicator is binned and presented by age groups. High blood pressure was included because of the potential linkage with heart disease and obesity. The obesity percentage for each neighborhood was estimated as the percent of adults who are overweight or obese as determined by BMIs based on aggregated individual surveys conducted from 1997 - 2001; rates ranged from 32% in Fenway to 62% in Mattapan (Figure 2).

Table 1. Information Potential for Health Indicator

Variables	Description	Unit
Heart	Heart disease mortality	Incidence rates (%)
Cancer	Cancer mortality	Incidence rates (%)*
Obese	Adults with overweight or obese	(%)
Hbp	Adults with high blood pressure	(%)

*The incidence rates are per 100,000 individuals

Our initial assessment is focusing on possible links between parameters typically available from census data and the various health indicators. Specifically, we are looking at possible linkages between obesity and ordinal census data such as age, ethnicity, and gender as well as the lack of exercise, educational attainment, median income, smoking rates, lack of health insurance, and poverty rate.

Table 2. Demographic information potentially related to obesity

Smoking	Adults who are current smokers	(%)
Nograd	Education rates of adults with less than high school	(%)

	diploma or GED	
Medinc	Median income	(%)
Insur	Adults who are uninsured	(%)
Poverty	Population below poverty rate	(%)
Educ	Education	(%)
lackexe	% of adults at risk of health problems due to lack of exercise	(%)

In addition to the more readily available census variables, we are considering a number of additional, less widely available parameters (Table 3) and their potential relation to the health indicators. One of the more interesting datalayers we recently compiled was the location of fast food suppliers and restaurants as a potential surrogate for dietary choices. Each restaurant was geo-coded based on its street address, compiled into a restaurant's datalayer (N=5,856), and the location then validated for consistency through verification of geocoded data points by neighborhoods in relation to the neighborhood address (Figure 2). Once located, restaurants were then subdivided into two categories (dine in or take out), and buffers were established around each location to assess accessibility to food shops. We are still exploring the relationships, but the notion is that restaurant accessibility and density, particularly to take out or fast food restaurants which by nature have higher fat contents, might help explain the observed obesity distribution.

Table 3. Spatial data layers potentially related to obesity

Res	Food Establishments	Points
Asp	After school programs	points
Childcc	Child care choices	points
Openspa	open space and recreational areas	polygons
Transp	Transportation choices (bus and subway)	Lines and points
Biket	bike trails	lines

Other useful spatial variables we have compiled and are considering include the location and amount of open space and recreational areas, educational facilities, bus routes, and subway lines. The availability and type of transportation may be correlated to physical exercise and activity which in turn may influence obesity rates. Educational knowledge regarding the risks and potential causes associated with obesity may be reflected in the spatial distribution for education and after school programs.

Data Analysis

With the dataset compiled, we have begun to analyze possible linkages of demographic and spatial environmental factors with observed human health distributions, particularly obesity, at a neighborhood scale. In this preliminary effort, we are looking at the spatial pattern relationships between health indicators and the various variables already mentioned (Tables 2 and 3). These selected variables were compared to health obesity rates visually as well as using a basic correlation analysis. Despite the limited sample size, multivariable linear regression analysis was also conducted to identify the rank of the importance among the variables. The correlation table as well as the visual analyses provides a means to begin to try to identify the potential spatial drivers and seek the appropriate spatial scales (neighborhood, census track, groups) for addressing and modeling the obesity problem.

RESULTS and DISCUSSION

Our preliminary visual review of the data layers did show some similarity in patterns between the health and demographic data. Figure 1 shows the patterns between the takeout restaurant distribution and the obesity rates at a neighborhood scale. In general, areas with a higher density of take out restaurants have a higher percentage of obesity, except for those neighborhoods that reside in the central downtown area (Back Bay, North End, South End and Allston Brighton). The disparity for these downtown areas may be due to the fact that restaurants in these areas are catering to visitors, students, or business people who do not typically reside in that neighborhood or may be related to the education level in the area seeing that 90% of the area's universities and colleges are found within these four neighborhoods.

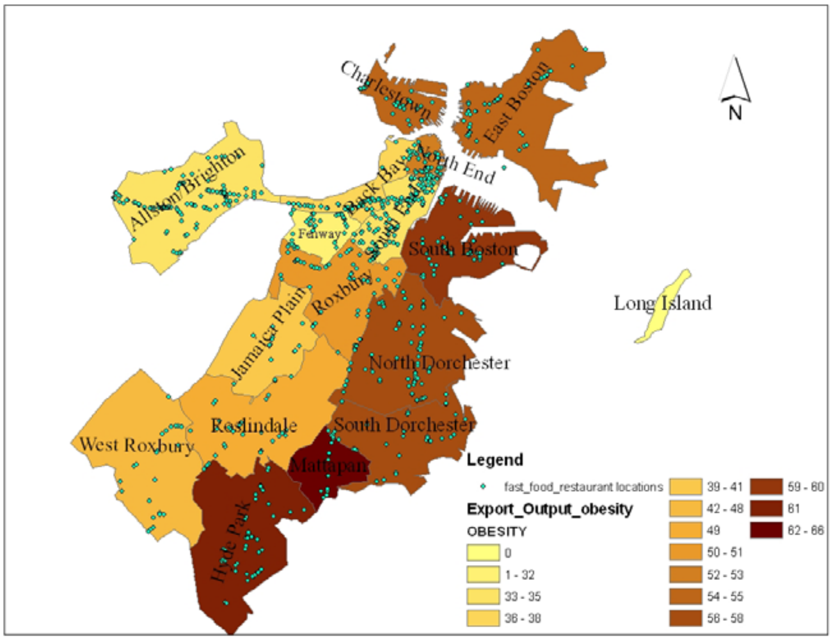

Figure 2. The spatial locations of takeout restaurants versus obesity rates in neighborhoods

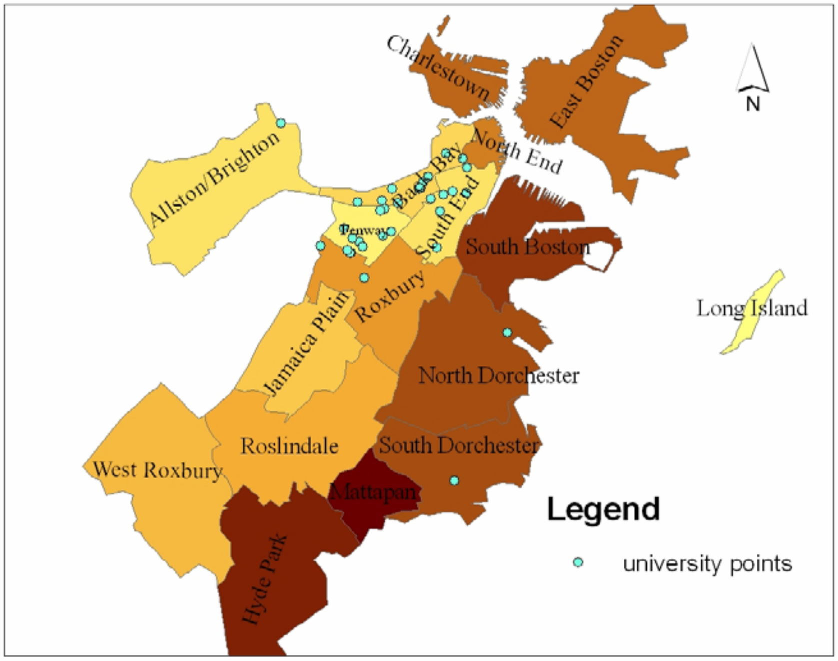

Figure 3. The locations of universities and colleges in Boston

The possible link between transportation and obesity is less compelling compared to the restaurant data. Figure 4 displays the location of subways and buses lines relative to the obesity distribution. In general, areas with greater access to public transportation exhibit lower rates of obesity. However, there are also several neighborhoods that do not match this result including West Roxbury, Roslindale, Mattapan, and Hyde Park. Overall, the link between transportation access and obesity is weak at best, suggesting that transportation grids alone are certainly not useful for quantifying obesity.

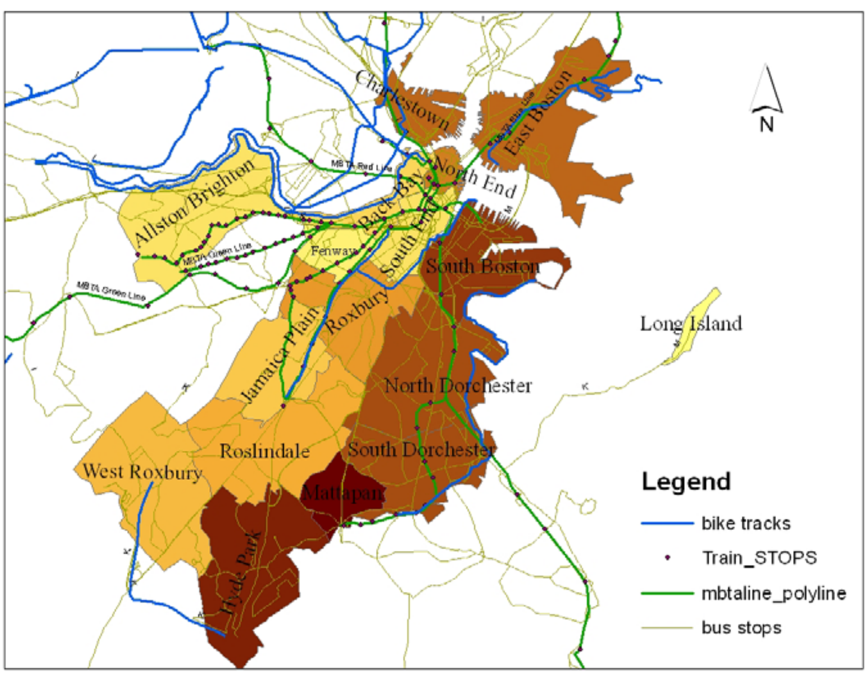

Figure 4. Obesity versus public transportation

A statistical analysis using the census data shows that obesity is at least partially correlated with income, ethnicity, and age. The results at the neighborhood scale are consistent with survey results conducted on an individual basis. The correlation between obesity and ethnicity may be related to genetics so care needs to be taken in identifying environmental risk factors that influence obesity from genetic factors.

Conclusion

Identifying the spatial drivers that influence obesity will be invaluable for helping fight the obesity epidemic. This preliminary analysis suggests that spatial patterns of less commonly available datasets like the location and density of restaurants might be able to partially explain the observed obesity distribution in Boston. Further study and consideration of additional parameters is clearly necessary, but the use of GIS opens the door for exploration of spatial linkages.

REFERENCES

Morland K, Wing S, Roux AD. (2002) The contextual effect of the local food environment on residents' diets: The atherosclerosis risk in communities study American Journal of Public Health 92 (11): 1761-1767

Shimakawa T, Sorlie P, Carpenter MA, Dennis B, Tell GS, Watson R, Williams OD. (1994) Dietary-intake patterns and sociodemographic factors in the atherosclerosis risk in communities study, Preventive Medicine 23 (6): 769-780

Shea S, Stein AD, Basch CE, Maylahn C, Strogatz DS, Novick L. (1991) Independent associations of educational-attainment and ethnicity with behavioral risk-factors for cardiovascular-disease, American Journal of Epidemiology 134 (6): 567-582

Morland K, Wing S, Roux AD, Poole C. (2002) Neighborhood characteristics associated with the location of food stores and food service places, American Journal of Preventive Medicine 22 (1): 23-29

Wu JL, Wang JF, Meng B, Pang LH, Song XM, Zhang KL, Zhang T, Zheng XY. (2004) Exploratory spatial data analysis for the identification of risk factors to birth defects, BMC PUBLIC HEALTH 4: Art. No. 23

Rothman KJ. (1990) A sobering start for the cluster busters conference, AMERICAN JOURNAL OF EPIDEMIOLOGY 132 (1): S6-S13 Suppl. S