Spatial Dynamics of the Local Food Environment in the City of Chicago: An Investigation of Data Sources and Methods

Michael D. Bader

Jennifer A. Ailshire

Population Studies Center
Institute for Social Research
University of Michigan

September 2007

Abstract:
Recent studies have demonstrated the importance of the local ecological environment on individual health outcomes. Data collection and analysis methods, however, need to be developed in order to extend our understanding about the mechanisms through which the local ecological environment contributes to individual health. In this paper we present a promising form of ecological data collection by combining commercial and observational data in addition to methodological innovations for analysis of these innovative data. We adapt the geostatistical method of kriging used in environmental and material sciences to demonstrate how the spatial autocorrelation of observations in the social environment can be used to derive accurate estimates of contextual variables from a relatively small sample of blocks in the city of Chicago. We use these estimates to measure the effect of the local ecological environment on individual body mass index. Implications for data collection and spatial analysis are discussed.

Introduction

Research on neighborhoods and health finds neighborhood disadvantage to be an important predictor of health outcomes including obesity (Robert and Riether 2004; Ellaway and Macintyre 1997), functional limitations (Balfour and Kaplan 2002), cardiovascular disease (Diez Roux 2003; Diez Roux et al. 2001) and all-cause mortality (Yen and Kaplan 1999; Bosma et al. 2001), and health behaviors such as diet (Diez Roux et al. 1999; Trout 1993) and physical activity (Yen and Kaplan, 1998; Ross 2000; Humpel, Owen, and Leslie 2002; Giles-Corti and Donovan 2002). Less work has been done to explore the specific mechanisms that link aspects of the neighborhood environment, such as poverty, to health outcomes (Diez Roux 2001). Living in disadvantaged neighborhoods is hypothesized to have an adverse effect on one's health because it means being exposed to the effects of concentrated poverty which includes disinvestment in neighborhood structures such as grocery stores and parks (Ellen, Mijanovich and Dillman 2001). However, lack of adequate data has prevented researchers from testing specific hypotheses concerning the effect of neighborhood structures on health.

In the absence of adequate data, aggregated measures of neighborhood disadvantage are used as 'global proxies' for the neighborhood built environment and thus do not capture actual features of the neighborhoods (Cummins et al. 2005a). Direct measurements of the local food environment are of particular importance because of the clear theoretical relationship to neighborhood disadvantage (Trout 1993) and health behaviors such as diet (Morland et al. 2002; McRory et al. 1999). Moreover, dietary intake is an important determinant of obesity and other chronic diseases. In fact, recent studies that have used innovative data collection techniques to obtain local ecological data have highlighted the importance of the food environment—meaning the availability of food stores in one's local community—on the health of residents (Cummins et al. 2005b; Morland et al. 2002).

Underlying these theoretical understandings of the local food environment is the notion of distance (i.e. proximity) to food stores. However, these studies have often relied on an implicit assumption of spatial relationships rather than explicitly modeling the spatial relationship to the food environment. Because they have used administrative boundaries of ecological units such as census tracts to proxy for spatial distance rather than modeling the actual spatial relationships themselves, important questions have remained unresolved. Among these is the theoretical and practical problem of identifying the spatial scale of the relationship between health outcomes and the local food environment that do not rely on *a priori* assumptions of neighborhood definitions.

Using a source of ecological data of the local food environment, we introduce a new analytical method that can overcome some of these obstacles and capture the spatial relationships of the built environment with health outcomes. Kriging is a family of methodological tools used widely in environmental and material sciences to estimate the presence of a physical process at a given location based on the spatial autocorrelation structure between sampled measurement locations. We describe how this method can be adapted for use in spatially dependent social and population sciences and specifically apply this method to the measurement of the local food environment. Specifically we use kriging to measure the scale and impact of spatial variation in the availability of food stores to explain differences in body mass index among residents in the city of Chicago and compare these methods to those currently used in the examination of neighborhood health effects.

Data and Measures

Food Environment Measures

The data of the food environment for this analysis comes from a database of employers compiled by the commercial proprietor, InfoUSA, and available for purchase by North American Industry Classification System (NAICS) codes. NAICS codes are detailed and standard industry classifications used in the United States, Canada and Mexico to facilitate economic and labor analysis by industry[1] and we focus specifically on restaurants in this analysis. The data are delivered in a business-level database with basic information including address, NAICS codes[2], number of employees and annual sales based on self-reports from employers.

We then geocoded each of these businesses based on their street address in the city of Chicago. Based on these geocoded addresses, we created counts of restaurants on each city block in Chicago. In turn, each block was geocoded to the geographical coordinates of its centroid. Therefore, this process provided us with a city block level dataset with the geographical location and count of the restaurants on each block in the city of Chicago.

Respondent Measures

The individual demographic and health measures are obtained from the Chicago Community Adult Health Survey (CCAHS). The CCAHS was administered between May 2001 and March 2003 to adults age 18 and older who lived in Chicago, IL. The survey is a stratified, multi-stage probability sample of 3,105 individuals with a response rate of 71.82%. The sample is stratified by neighborhood, defined as an area with varying numbers of adjacent city clocks, with several 5-10 respondents residing in each neighborhood cluster. The CCAHS sample was drawn from 343 neighborhood clusters (NCs) with over-sampling in 80 NC's that represent a socioeconomically and racially/ethnically heterogeneous subset of Chicago's neighborhoods. The CCAHS provides indicators of individual health outcomes, health behaviors, and sociodemographics. Body mass index (BMI) is calculated using self reports of height and weight. The CCAHS survey also contains sociodemographic measures such as age, gender, race/ethnicity, nativity, income, education, employment status, and marital status.

Systematic Social Observation of Sampled Blocks

In addition to the individual level survey, the CCAHS concurrently conducted a survey of the ecological conditions on the block surrounding each respondent's house. This survey used the systematic social observation (SSO) method to observe the environment and look for specific items. The SSO method uses trained raters to indicate and measure specific items in a sampled area such as a street or a block. Because the responses are systematically collected and raters are trained to observe the same items, the measurements are comparable across sampled areas and are not unlike systematic survey responses (Reiss 1975). This "ecometric" approach allows researchers to obtain measurements of respondents' physical and social surroundings that are independent of the respondents, thus allowing for statistical analysis of respondents' surroundings (Raudenbush and Sampson 1999; Sampson and Raudenbush 1999).

---

[1] http://www.census.gov/epcd/www/naicsdev.htm

[2] Every employer could designate up to four NAICS codes, though each was asked to indicate their "primary" NAICS code.

The measurements obtained from the SSO method can be much more detailed than obtaining data from administrative or commercial databases, such as that acquired form InfoUSA. However, this increased depth comes at the price that it is much more costly to collect; therefore, it is necessary to create a sample of blocks rather than attempting to obtain this data on every block. The decision was made in the CCAHS to collect SSO data on the block immediately surrounding each respondent's home and, thus, the sample is drawn to a population of non-institutional residences rather than the population of all blocks in the city.

Strategy and Methods

Analytic Strategy

The goal of this analysis is to determine whether the use of kriging can improve the measurement of the local food environment compared to methods that are commonly applied in current demographic research when measuring spatially autocorrelated *sample* data. Obviously the best strategy for population research when complete data is available for an independent variable is to use that data; and, in fact, recent analyses of health outcomes have used spatial methods with such measures (Diez Roux, et al. 2007). However, it is often not possible to obtain complete data on independent variables of interest and researchers are then required to rely on samples of those data. The more control over the process of measurement that researchers exert or the more detail desired in the measurement of the environment requires that non-routine and innovative methods be applied. In particular, many items that are—or could be—obtained from the SSO method require that sampling be conducted.

With data from both the commercial database obtained from InfoUSA and the sampled blocks from the SSO, we have the advantage of having both the complete data of restaurants in the city of Chicago at the same time that we have a sampling strategy employed for use with a SSO. Therefore, our overall strategy is to "sample" the complete data from the InfoUSA database by creating a separate dataset containing only those blocks that were part of the CCAHS SSO sample. We will then compare our estimates of the restaurant or bar environment derived from the application of kriging within neighborhoods with the actual restaurant or bar environment measured by the complete InfoUSA data. Finally, we will use the kriged estimates to whether we are able to obtain comparable estimates to the complete InfoUSA data for the effect of neighborhood food environment on BMI.

Statistical Methods

In order to derive estimates of the restaurant environment in the city of Chicago, we use *universal kriging* to estimate the number of restaurants,that exist on the non-sampled blocks in the city of Chicago. This is two part process. In the first step we fit a *semivariogram* which is, essentially, a function that describes the variance in the outcome as a function of the separation distance between two sampled observations. This function defines the variance that will be used to estimate the weights used in the kriging function to derive the estimates of the outcome at unsampled locations (see Figure 1 for the initial variogram used for the estimation of restaurants).

Second we estimate the value of the outcome at unknown locations. Based on the spatial variance structure derived from the semivariogram, we estimate the outcome at unknown locations through the use of *universal kriging*. Universal kriging is a method which uses the spatial dependence in the known, sampled locations and fits weights to each of those observations to estimate the presence of the outcome in an unsampled location. The measurement of the outcome at each of the sampled locations is then multiplied by the weight attributed to that sampled location based on both its distance from the unsampled location we wish to estimate and its distance from other sampled locations. In this way, universal kriging controls for both the spatial dependence between the unsampled and sampled locations as well as between all of the sampled locations. Therefore, each of the weights given to a sampled location in a geographical cluster is reduced to account for the redundancy among proximate sampled locations while geographically isolated sampled locations are given greater weight, particularly if they are close to the unsampled location we wish to estimate.

Additionally, with universal kriging we can also account for the mean level of variation based on other characteristics of the block. For instance, because we expect to find restaurants more frequently on busy commercial thoroughfares rather than small, exclusively residential streets, we can add a variable in the system of kriging equations that accounts for this local variation in the mean. We can also account for error in our estimates and, using the system of kriging equations, also place confidence bounds around our estimates at the unsampled locations.

Both the estimates and the confidence intervals can also be used to predict the value of an areal unit as well as a specific point location. By using this method to derive estimates for levels of analysis such as block groups, census tracts or neighborhood clusters, our results can be combined with other methods of analysis such as hierarchical linear models that have been extensively used in the study of neighborhoods and health (Diez Roux 2001). We will use these kriged estimates of the presence of restaurants across neighborhood clusters (NCs) and measure it against the actual presence of restaurants from the complete InfoUSA dataset. Furthermore, we will also compare how well the kriged estimates perform in estimating an accurate picture of neighborhood compared to the standard method of simply taking the average value across sampled blocks within a neighborhood cluster.

Finally, using hierarchical linear models we will predict individual-level BMI using all three measures of restaurant presence (i.e. calculations based on the complete InfoUSA dataset, our NC-level kriged estimates and the simple mean across sampled blocks). Performing this analysis will indicate whether using different measures of the neighborhood-level food environment change our understanding of neighborhood contextual factors on BMI.

Preliminary Findings

Initial analysis indicates that the sample of SSO blocks, which is available for only a small sample of the city blocks, can be used to estimate values for the larger urban areas. Figure 2 shows an example of kriged estimates for the presence of restaurants on city blocks. On this map, the chloropleth areas indicate where it is more and less likely to find a restaurant with lighter colors showing where we would expect a lower probability of finding a restaurant and darker areas indicating higher probability. In addition, we map the locations of all of the restaurants taken from the entire InfoUSA dataset as well as the location of major roads.

We only show a small snapshot of the city of Chicago in order to protect the confidentiality of respondents[3]. From this snapshot, one can see that we do a reasonably good job predicting where restaurants are present. We can also see that controlling for whether the block sits on a major road (indicated by grey, yellow or red lines) increases the estimate on that block and that, this seems to be an effective strategy for predicting the location of restaurants in Chicago.

Contributions of Research and Further Steps

As we mentioned, there is no reason to use the kriged estimates of the sampled data when we have the actual complete dataset. However, being able to demonstrate the effectiveness of this strategy by being able to compare the estimates derived from kriging to the complete data available from the InfoUSA dataset may show how this method might be effective for other measures and outcomes. For instance, if we discover that the amount of physical disorder found on blocks is spatially autocorrelated (which, we believe that it is), this method of estimation could enable us to derive better estimates of physical disorder within neighborhoods or tracts. Furthermore, collecting the kind of complete data that we were able to collect for the city of Chicago might not be possible for studies with a larger sampling frame. Kriging could provide a way in which ecological measurement can be cost-effectively produced in national samples such as NHANES, HRS, PSID, etc.

Of course, in this analysis, we advocate for understanding the spatial scale and variance structure of data but only apply this to the independent variables in the analysis. In the end, we end up using NCs to analyze the relationships between the restaurant environment and BMI. It will also be important to extend this method to analyze the dependent as well as the independent variables to account for spatial scale and spatial covariance between dependent and independent variables. This does not mean that there are not good reasons for using areal definitions of neighborhoods. Research has demonstrated that ecological boundaries are important to the development of neighborhood patterns (Grannis 1998, 2005). This research does raise questions about the mechanisms and theoretical underpinnings of neighborhood health and population research. Developing testable hypotheses about the spatial relationships of ecological conditions and their relationship to population outcomes requires that researchers think about whether areal or distance-based measures best describe the process under investigation.

[3] Because one of the properties of the system of kriging equations is that it perfectly interpolates values at the sample locations; therefore, displaying the entire map of Chicago would indicate the sampled blocks and revealing the location of our respondents.

References

Balfour, J.L. and Kaplan, G.A. 2002. "Neighborhood environment and loss of physical function in older adults: evidence from the Alameda County Study." American Journal of Epidemiology 155:507-515.

Bosma, H., Dike van de Mhenn, H., Borsboom, G.J.J.M., and Mackenbach, J.P. 2001. "Neighborhood socioeconomic status and all-cause mortality." American Journal of Epidemiology 153: 363-71.

Cummins, S., Macintyre, S. Davidson, S., and Ellaway, A. 2005a. "Measuring neighborhood social and material context: generation and interpretation of ecological data from routine and non-routine sources." Health and Place 11 249-260.

Cummins, S., Stafford, M., Macintyre, S., Marmot, M., and Ellaway, A. 2005b. "Neighborhood environment and its association with self rated health: evidence from Scotland and England." Journal of Epidemiology and Community Health 59:207-13.

Diez Roux, A. V. 2003. "Residential environments and cardiovascular risk." Journal of Urban Health 80:569-589.

——————. 2001. "Investigating neighborhood and area effects in health." American Journal of Public Health 91:1783-1789.

Diez Roux, A.V., Evenson, K.R. McGinn, A.P., Brown, D.G., et al. 2007. "Availability of Recreational Resources and Physical Activity in Adults." American Journal of Public Health, 97: 493-499.

Diez Roux, Ana V., Nieto, F.J., Caulfield, L., Tyroler, H.A., Watson, R.L., and Szklo, M. 1999. "Neighborhood differences in diet: the Atherosclerosis Risk in Communities (ARIC) Study." Journal of Epidemiological Community Health 53:55-63.

Diez Roux, A.V., Stein Merkin, S., Arnett, D., Chambless, L., Massing, M., Nieto,F.J., Sorlie, P., Szklo, M., Tyroler, H., and Watson, R.L. 2001. "Neighborhood of residence and incidence of coronary heart disease." New England Journal of Medicine 345: 99-106.

Ellaway, A., Anderson, A., and Macintyre, S. 1997. "Does area of residence affect body size and shape?" International Journal of Obesity Related Metabolic Disorder 21:304-308.

Ellen, I.G., Mijanovich, T., and Dilliman, K-N. 2001. "Neighborhood effects on health: exploring the links and assessing the evidence." Journal of Urban Planning 23: 391-408.

Giles-Corti, B. and Donovan, R.J. 2002. "Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment." Preventative Medicine 35:601-611.

Grannis, Rick. 1998. "The Importance of Trivial Streets: Residential Streets and Residential Segregation." American Journal of Sociology 103:1530-64.

——————.2005. "T-Communities: Pedestrian Street Networks and Residential Segregation in Chicago, Los Angeles, and New York." City & Community 4:295-321.

Humpel, N., Owen, N., and Leslie, E. 2002. "Environmental factors associated with adults' participation in physical activity: a review." American Journal of Preventative Medicine 22:188-199.

McCrory, M.A., Fuss, P.J., Hays, N.P., Vinken, A.G., Greenberg, A.S., and Roberts, S.B.1999. "Overeating in America: association between restaurant food consumption and body fatness in healthy adult men and women ages 19-80." Obesity Research 7:564-571.

Morland, K., Wing, S., Diez Roux, A. and Poole, C. 2002. "Neighborhood characteristics associated with the location of food stores and food service places." American Journal of preventative Medicince 22(1) 23-29.

Raudenbush, S.W., and Sampson, R.J. 1999. "Ecometrics: Towards a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods." *Sociological Methods* 29:1-41.

Reiss, A.J., Jr. 1975. "Systematic observation surveys of natural social phenomena." Perspectives on attitude assessment: surveys and their alternatives. NTIS, 132-150.

Robert, S.A., and Reither, E.N. 2004. "A multilevel analysis of race, community disadvantage, and body mass index among adults in the US." Social Science and Medicine 59: 2421-2434.

Ross, C.E. 2000. "Walking, exercising, and smoking: does neighborhood matter?" Social Science and Medicine 51:265-74.

Sampson, R.J. and Raudenbush, S.W. 1999. "Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods." *The American Journal of Sociology* 105: 603-51.

Troutt D. 1993. The Thin Red Line: How the Poor Still Pay More. San Francisco, CA: Consumers Union, West Coast Regional Office; 1993.

Yen, I.H. and Kaplan, G.A. 1998. "Poverty area residence and changes in physical activity level: evidence from the Alameda County Study." American Journal of Public Health 88:1709-12.

Yen, I.H., and Kaplan, G.A. 1999. "Neighborhood and social environment and risk of death: multilevel evidence from the Alameda County Study." American Journal of Epidemiology 149:898-907.

Figure 1. Semivariogram of Restaurants from Sample of Chicago Blocks

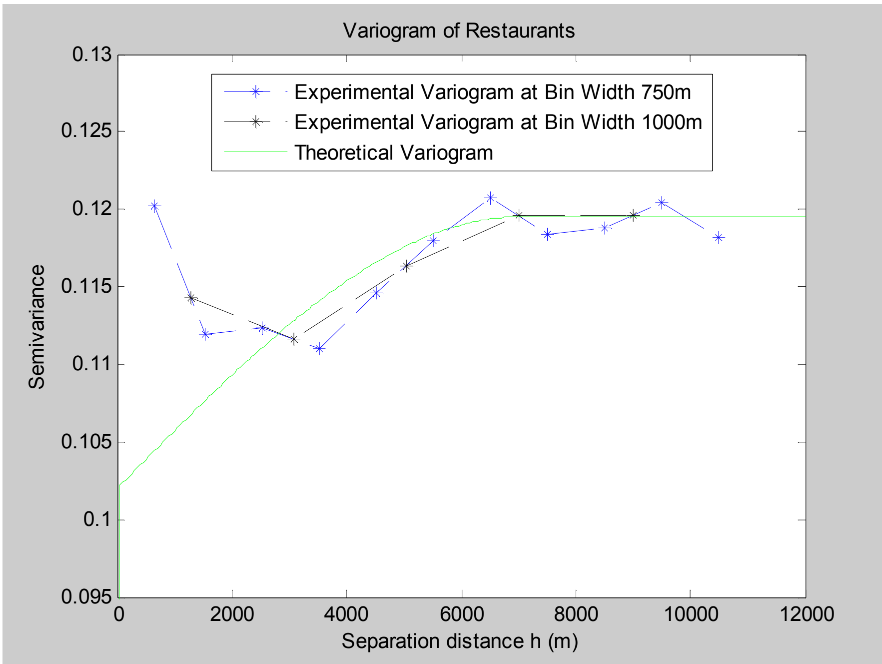
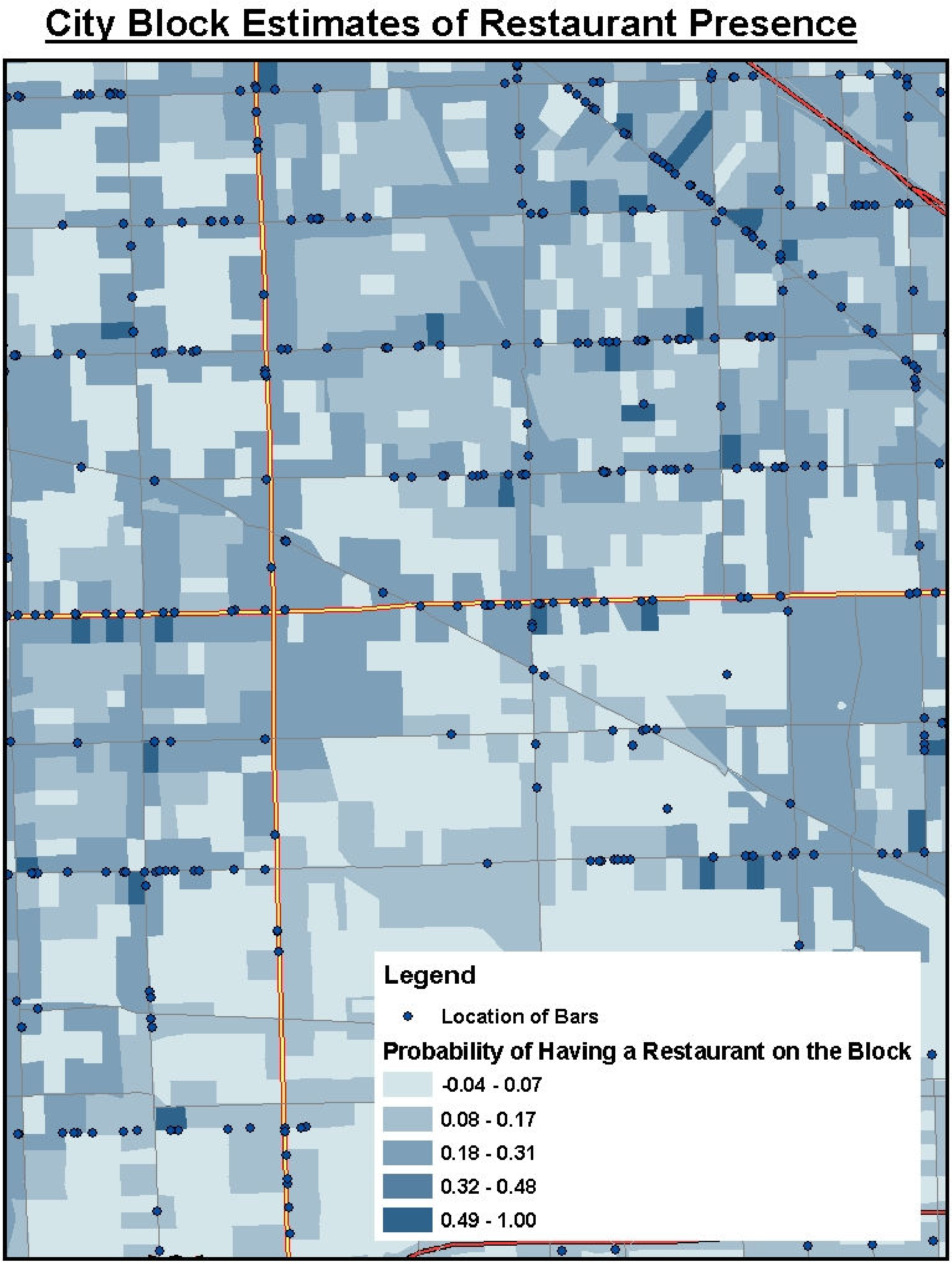


Figure 2. Kriged Estimates of the Presence of Restaurants, City of Chicago



Note: Full extent of Chicago not shown in order to protect confidentiality of respondents