Title
The Individual Association Between Food Store Types and Body Mass Index in Los Angeles County

Author
Capone-Newton, Peter

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Individual Association Between Food Store Types and Body Mass Index in Los Angeles County

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Urban Planning

by

Peter Capone-Newton

2013

ABSTRACT OF THE DISSERTATION

The Individual Association Between Food Store Types and Body Mass Index in Los Angeles County

by

Peter Capone-Newton

Doctor of Philosophy in Urban Planning

University of California, Los Angeles, 2013

Professor Paul M. Ong, Chair

Using the Los Angeles Family and Neighborhood Survey (L.A. FANS), detailed individual-level data on shopping location, store name, and body mass index are analyzed to assess relationships between body mass index and food store types. The analysis groups similar store brands to create unique food store types, providing finer discrimination than industrial classification or annual sales volume. Seven food store types are created: English-language major supermarket chains, discount food stores with “less”, “save”, or “bargain” in the name, Spanish-language name supermarkets or grocery stores, specialty stores defined as having fewer locations, smaller format, specific product focus, and/or limited product inventory, and independent, small, and bulk food stores. Documentation within the dataset of shopping locations, home locations, and car ownership allow strict control for distance and transportation. Accounting for these and other individual and neighborhood characteristics using multivariate regression models, body mass index is significantly lower in people who shop in specialty and Spanish-language name food store types compared to major chain food store types in higher poverty neighborhoods. Additional analysis indicates that reported supermarket shopping rates are higher than expected based on the prevalence of supermarkets in home Census tracts. Absence of neighborhood supermarkets (in home Census tracts) is a common state among respondents across all poverty strata, although more common in very poor tracts. In a subgroup of non-movers over a six-year period, opening and closing of stores is associated with change in shopped store type. Because the main results are cross-sectional, causal inference is difficult. Store types may influence body mass index, or individuals may have unobserved characteristics, which explain the association between store types and body mass index. Further research is needed to assess the direction of association. However, these results suggest specific store types should be the focus of policy and research rather than broader categories defined by sales volume or industrial classification. Absence of supermarkets may be an insufficient tool to characterize shopping behavior. Store opening and closing may stimulate change in store type preferences, but whether this change in store type is associated with change in health behaviors or health outcomes is unknown.

The dissertation of Peter Capone-Newton is approved.

Evelyn A. Blumenberg

Lois M. Takahashi

Arleen F. Brown

Paul M. Ong, Committee Chair

University of California, Los Angeles

2013

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This work was funded in part by training grants provided by the National Institutes of Health (NIH) National Research Service Award (NRSA), the Specialty Training and Advanced Research (STAR) program of the Department of Medicine, University of California Los Angeles (UCLA), the Department of Urban Planning at UCLA, and the American Heart Association (AHA) Pharmaceutical Round Table/Spina family.

Manuscripts based on this work are in preparation for publication, co-authored with Arleen F. Brown and Paul M. Ong.

Map data are provided by Google and other data providers licensed by Google as indicated on maps contained in this document and are used under the terms provided for print use as defined in http://www.google.com/permissions/geoguidelines.html (Accessed May 3, 2013) and https://developers.google.com/maps/terms (Accessed May 3, 2013).

Food store data were purchased from InfoUSA and used under the terms and conditions provided by InfoUSA at http://www.infousa.com/terms-conditions/ (Accessed May 6, 2013).

BIOGRAPHICAL SKETCH

Peter Capone-Newton M.D., M.P.H., M.A. is an American Heart Association (AHA) Pharmaceutical Round Table (PRT) and Spina Family Outcomes Research Fellow in Department of Neurology in the Geffen School of Medicine at the University of California Los Angeles (UCLA). Additionally he is a Specialty Training and Advanced Research (STAR) Fellow in the Division of General Internal Medicine and Health Services Research within the Department of Medicine. He is board certified in Public Health and General Preventive Medicine.

Prior to his current fellowship positions Dr. Capone-Newton was a National Research Service Award (NRSA) fellow at UCLA. He completed his residency training in Preventive Medicine within a program sponsored jointly by the Los Angeles County Department of Public Health and the California Department of Public Health. During that time he worked within the newly formed Policies for Livable Active Communities and Environments (PLACE) program at the department of public health, an innovative program designed to apply built environment interventions to public health problems. In his first year of residency training he was a preliminary medicine intern at Olive View-UCLA Medical Center, a Los Angeles County Department of Health Services hospital in Sylmar, CA.

Prior to residency training, Dr. Capone-Newton was Associate Director of Life Science Solutions at Cerner Health Insights (formerly Zynx Health). In that position he coordinated content development for a web-based relational database of medical literature, working daily with content experts, database architects, and web designers.

Dr. Capone-Newton currently holds a Masters of Arts (MA) degree in Urban Planning from the University of California Los Angeles. He completed his medical and public health degrees in the combined MD/MPH program at the Tufts University School of Medicine in Boston, MA. He completed a BS in Microbiology and Immunology from McGill University, in Montreal, Quebec, Canada, as well as a minor in the Social Studies of Medicine.

Chapter 1 - Introduction

The disciplines of urban planning and public health emerged from each other over a century ago, since diverging into the modern fields practiced today. Urban planning in the Victorian era was created to alleviate the health ailments of overcrowding and filth, while at the same time, physicians like John Snow began to document the systematic patterns of disease caused by poor sanitation and drinking water delivery, developing the science of public health, epidemiology (Hall 2002; Krieger 2011). Thus it is fitting, that in the past decade or so, the two have become reacquainted with their common beginning (Frumkin, Frank, and Jackson 2004). This research is an example of such scholarship, returning to the common themes of urban planning and public health origins.

That the two come together at this time is not surprising. Both disciplines face challenges unique to the twenty-first century that neither faced in the twentieth. A majority of the world's population now live in cities and future population growth is expected in urbanized areas, making urban planning more relevant than ever (United Nations 2012). Chronic health conditions such as heart disease and cancer, developed slowly over time, now are the leading causes of death globally as opposed to acute communicable diseases, necessitating a rethinking of old public health and medical approaches to disease care (Yach et al. 2004). Public health professionals are beginning to seek remedies for the new burden of chronic conditions, which develop over time, reflecting the cumulative effect of numerous individual decisions and social interactions, requiring a new understanding of how those social interactions develop and can be moderated to promote health. In turn, urban planners must plan for urbanizing populations, accommodate new planning-related demands on the profession like the integration of public health interests into planning decisions, and contribute to understanding how spatial interactions affect individual and population outcomes. These new realities mean that both urban planners and public health professionals have to reconsider how the health of populations and individuals are affected by the physical organization and social relationships of the cities they inhabit. This research provides insight for both disciplines.

As exemplified in the historical links between urban planning and public health, the connection between the two has often been based upon human development effects on the natural environment, specifically water and air quality. For example, many transportation planning interventions are implemented specifically in response to the negative health effects of air pollution (Wachs 1993). This research diverges from that tradition by considering the long-term health effects of food provision by the private market.[1] The importance of food provision by the private market has extended to long-term effects from the typical short-term focus on food safety precisely because of the changing nature of diet-related disease.

---

[1] One likely reason for the focus on air and water, versus food, is because of the class of economic good each represents. Air and water are public and common goods respectively, and food is a private good, so the role of government intervention in its provision is conceived of differently.

The goal of this research is to better understand how spatially-based behavior helps explain the relationship between food shopping, store choice, and health as measured by body mass index (BMI). The following are the three key research questions, whose answers may provide new insights into the relationship between these spatial behaviors and health outcomes:

1) How similar are supermarket measures derived from ecologic imputation[2] with reported behavior?[3]
2) Are food store types[4] associated with body mass index (BMI)?
3) Are store closures/openings associated with store type changes over 6-year follow-up?

This research focuses on chain supermarkets as a crucial element of the food environment because health researchers have hypothesized they contain sufficient food variety to facilitate consumption of a healthy diet compared to other stores (Sallis et al. 1986). Their hypothesized positive influence on diet and body weight is supported empirically in the public health literature.[5] These diet and body weight outcomes, combined with the unequal spatial distribution of chain supermarkets observed in some places (Morland et al. 2002), may

---

[2] The observed prevalence of supermarkets in home Census tracts.
[3] Based on survey responses.
[4] Store types are groupings of store brands conceived to be close substitutes.
[5] A review of this literature is contained in analytical chapters (3-5) and summarized in appendix C.

contribute to observed health disparities in specific income and racial/ethnic groups.

Data

All three of the analytical chapters in this dissertation come from a common core dataset constructed from three sources: the Los Angeles Family and Neighborhood Survey, the United States Census Bureau, and InfoUSA, a private data provider of business listings. This section discusses these sources, reducing the need to discuss them repeatedly in each of the analytical chapters. The individual analytical chapters discuss any supplementary data or chapter specific data issues.

The Los Angeles Family and Neighborhood Survey, known as L.A. FANS, is a multi-disciplinary, longitudinal social science survey conducted via in-person interview. The survey was specifically designed to address the role of neighborhoods in the outcomes of children, adults and families, reflecting the nature of research interests over the past twenty years, on the potential role of these contexts on individual and family outcomes. It is similar in approach to the Project on Human Development in Chicago Neighborhoods, known as PHDCN, another large social science survey designed to assess the influence of neighborhoods on individual outcomes (Sastry et al. 2006).

As the principal investigators Narayan Sastry and Anne Pebley note, the theoretical framework for the L.A. FANS study is based on the broad hypothesis that multiple social structures (or “environments”) such as households, work, school, social networks, and other social institutions affect a range of individual outcomes. In addition to these social structures, residential neighborhoods themselves have been specifically hypothesized to play an additional role in individual outcomes and are the focus of L.A. FANS. This hypothesis comes from both the early twentieth century Chicago school sociologists[6] who described the structure and change of neighborhoods, and more recent research that has focused on concepts like collective efficacy, social capital and the role of concentrated poverty and racial segregation in reinforcing social norms[7] (Sastry et al. 2006).

Despite the difficulty in defining the geographic boundaries of a neighborhood, L.A. FANS researchers determined that Census tracts would comprise the primary sampling unit of the survey. Within this unit, level of poverty (very poor, poor and not poor[8]) was a measure used for sample stratification. In addition, households with children were oversampled. This approach permits more detailed inference for families, and comparison of outcomes across levels of poverty. Both levels of analysis were hypothesized to influence individual outcomes, in addition to the local social structure defined by Census tracts.

The survey was conducted at two time points approximately six years apart, 2001-2002, and 2007-2008. For the first analytical chapter (Chapter 3),

---

[6] Sastry and Pebley cite Burgess, Park and Hawley.
[7] Sastry and Pebley cite Sampson, Coleman, Wilson, Massey and Denton.
[8] See appendix A describing L.A. FANS poverty definitions.

data from wave 1 (2001-2002) are used, comprising 2297 adult respondents. In the second analytical chapter (Chapter 4), data from wave 2 (2007-2008) are used, resulting in 915 adult responses. In the final analytical chapter (Chapter 5), wave 2 data from non-movers are used, resulting in a sample of 620 adults.[9]

The focus of analysis in this work is based upon the question: "What store do you (and others in this household, if more than one adult in the household) normally go to buy groceries?"[10] The respondent was asked to report the name of a single store and cross streets for that store. If the respondent attempted more than a single response then the interviewer asked, "What is the place you generally get most of your groceries?"[11] In datasets available to outside researchers, L.A. FANS research staff supplied the reported store name and a geocode (latitude/longitude) based upon the cross streets reported. The research dataset also included a geocode quality flag indicating the quality of match between reported cross streets and known intersections as determined by the geographic information system in use by staff.[12]

The primary outcome, body mass index, is also provided by L.A. FANS. In wave 1 this outcome was self-reported, based upon self-reported height and

---

[9] A detailed accounting of the sample selection and comparison of wave 1 and wave 2 main samples is provided in appendix A.

[10] This question does not specify a geographic context for the question; thus while the survey focuses on neighborhoods, the context of shopping as defined by this question is not limited to this context.

[11] These are survey questions [A]B15 (store name) and [A]B16 (cross streets) in the Adult Questionnaire available at http://lasurvey.rand.org/documentation/questionnaires/request.html

[12] This is question AB16FG in research datasets and the L.A. FANS wave 1 codebook (RAND publication DRU-2400/2-1-LAFANS); Available at http://www.rand.org/pubs/drafts/DRU2400z2-1.html

weight.[13] In wave 2 BMI was both self-reported and directly measured by survey staff.[14] The main analysis contained in chapter 4 uses directly measured BMI. Additional BMI-related questions relevant for this analysis were added in the wave 2 survey. These include: fruit and vegetable consumption, fast food consumption, and physical activity.[15]

Poverty estimates for the year 1997 in 1990 Census geography, and change in poverty from 2000 (Decennial Census, 2000) to 2009 (American Community Survey, 2005-2009 estimates) in 2000 Census geography were used in the analysis. The 1997 Census poverty estimates were provided by the RAND corporation via the Inter-university Consortium for Political and Social Research (ISCPR) (Escarce, Lurie, and Jewell 2011). Year 2000 and 2005-2009 poverty estimates were provided by direct download from the U.S. Census. Year 1990 detailed (versus cartographic) Census geographies based on original Census data were directly downloaded from the National Historical Geographic Information System (NHGIS) at the University of Minnesota.[16] Year 2000 detailed Census geographies were provided by direct download from the U.S. Census.

---

[13] These are survey questions [A]M15 (weight) and [A]M16 (height) in the Adult Questionnaire reported in English or metric system units depending on the respondent preference. BMI was calculated by the author based on the equation, BMI = weight (kg) / height squared (m^2)

[14] BMI was provided pre-calculated in L.A. FANS datasets.

[15] Differences between wave 1 and wave 2 availability of relevant questions for this analysis are summarized in appendix A.

[16] https://www.nhgis.org

Business locations were purchased from InfoUSA for the years 2003 and 2009. InfoUSA[17] provides business lists with data including industrial classification based on the North American Industrial Classification System (NAICS) and estimated sales and employee counts. Street addresses were included in the data, as are latitude and longitude with a geocode quality flag. The data were collected using a proprietary data collection system that includes use of existing directories and other secondary data sources combined with independent data verification of each listing.[18]

Street network data were obtained from Environmental Systems Research Institute (ESRI) included with ArcGIS 10 software. The dataset, "North America Detailed Streets" contains "detailed streets, interstate highways, and major roads within the United States and Canada."[19] According to the data description provided, data were from the year 2007 and originate from TomTom[20] and ESRI.

---

17 http://www.infousa.com

18 According to InfoUSA, "We gather our information from a multitude of directory and event-driven sources, including new business filings, daily utility connections, press releases, corporate websites, annual reports, user-generated feedback, and thousands of U.S. and Canadian Yellow Page directories. And because we maintain an intimate knowledge of our sources and complete control over our compilation processes, we're able to continually improve our methods to ensure the best data possible. Then, we do something no other data provider does. We call each and every business—making over 40 million calls each year—to gather and verify valuable information and ensure your data is current, accurate, and relevant." From: http://www.infousa.com/business-lists/ (Accessed: May 20, 2013)

19 http://www.arcgis.com/home/item.html?id=f38b87cc295541fb88513d1ed7cec9fd (Accessed: May 20, 2013)

20 http://www.tomtom.com/en_us/

Methods

Data from L.A. FANS, the U.S. Census, and InfoUSA are spatially linked using geographic information systems (GIS). Two separate GIS processes are performed based on whether the data linking took place within the secure data enclave (SDE) housing L.A. FANS data or using publically displayable data.

L.A. FANS provided latitude and longitude of home address and store cross streets for respondents. Using geocodes provided by InfoUSA, store responses from each wave are matched to stores in the InfoUSA database using ArcGIS 10. Store locations are joined to U.S. Census tracts in shapefile format using ArcGIS 10. Distances between home and store are calculated for the shortest distance street network path. Public facing visualizations of InfoUSA data are created using Google Fusion Tables and the Google Maps Application Programming Interface (API). Detailed methods are described in each analysis section.

Store name responses are grouped into common store types. Store types are created based on both deductive and inductive reasoning. Store types provide more aggregation than simply using store brands alone, acknowledging that some brands have common characteristics and are generally considered close substitutes. Store types are less aggregated than simply grouping all stores by NAICS classification or based on sales volume or other size measures.

A single dataset for statistical analysis is created from the GIS linked datasets. In chapter three descriptive statistics and sensitivity and specificity calculations are performed. In chapter four, body mass index is the dependent variable and store type is the main independent variable of interest. For these analyses, typically multilevel regression models[21] are used to account for clustering of observations by Census tract and permit interference at that unit of observation. In chapter five, store type change is the dependent variable and store openings/closings is the main independent variable.

The data are limited in several ways. First, the survey question limits store responses to a single response. Store shopping frequency and travel mode to store are unknown. Store purchases and store contents are unknown.

Summary of Findings

The findings in these analytical chapters are largely consistent with *a priori* assumptions. In chapter three, I use a standard definition of supermarket chain based on industrial classification, and find that presence of a supermarket in a home census tract is less frequent than reported shopping in a supermarket. Most respondents live in Census tracts without supermarkets in their home tract, yet a high proportion of respondents in those tracts report shopping in supermarkets. For example, for residents of poor tracts as defined by L.A. FANS, 15% of respondents live in Census tracts with supermarkets, but 58% of these respondents shop in supermarkets. None of the 20 very poor sampled Census tracts contain supermarkets, but 38% of respondents shop in

---

[21] Multilevel model characteristics and functional format are described in Appendix F "Rationale for Statistical Methods."

supermarkets. The absence of supermarkets in very poor Census tracts compared to poor and non-poor tracts is a finding similar to early work in this literature and highlights potential differences in proximity to supermarkets across poverty strata.

In chapter four, I introduce the concept of store types. Brands are grouped into seven mutually exclusive types: major chain, discount, specialty, Spanish-language, independent, bulk and small. Based on this novel grouping of stores, multivariate regression models estimate the association between directly measured body mass index and store type accounting for individual, household, and Census area-level characteristics. These characteristics include individual fast food consumption and physical activity, as well as street network distance to the store and household car ownership. The analysis shows that specialty stores and Spanish-language stores are associated with similarly lower body mass index compared with major chain stores in higher poverty tracts. These results occur despite the geographically distinct areas these two store types occupy.

In chapter five, I characterize six-year inter-temporal change in store type along with factors associated with this change. Over a six-year period most respondents shop in stores present at both time periods with a low rate of store type change. However, among respondents who experience some change in the stores available to them over this period of time, either stores opening or closing, the rate of store type change is statistically significantly higher, after accounting for individual characteristics and change in individual characteristics over time.

The final chapter of the dissertation discusses the results and implications of the research. In the first analytical chapter I show the limitations of aggregate imputation of shopping behavior suggesting that more detailed measurement of shopping behavior might assist further research on the association between shopping behavior and health outcomes. In the second analytical chapter I find an association between specific store types and lower body mass index. While the research is cross-sectional, it suggests there may be a role for promotion of specific store types, and that store type may be an appropriate method for store classification. In the third analytical chapter I find that store change over a 6-year period is associated with store type change, suggesting that policies which promote store openings or closing may stimulate change in the type of store shopped in by someone in the vicinity of this change.

Chapter 2 - Conceptual Background

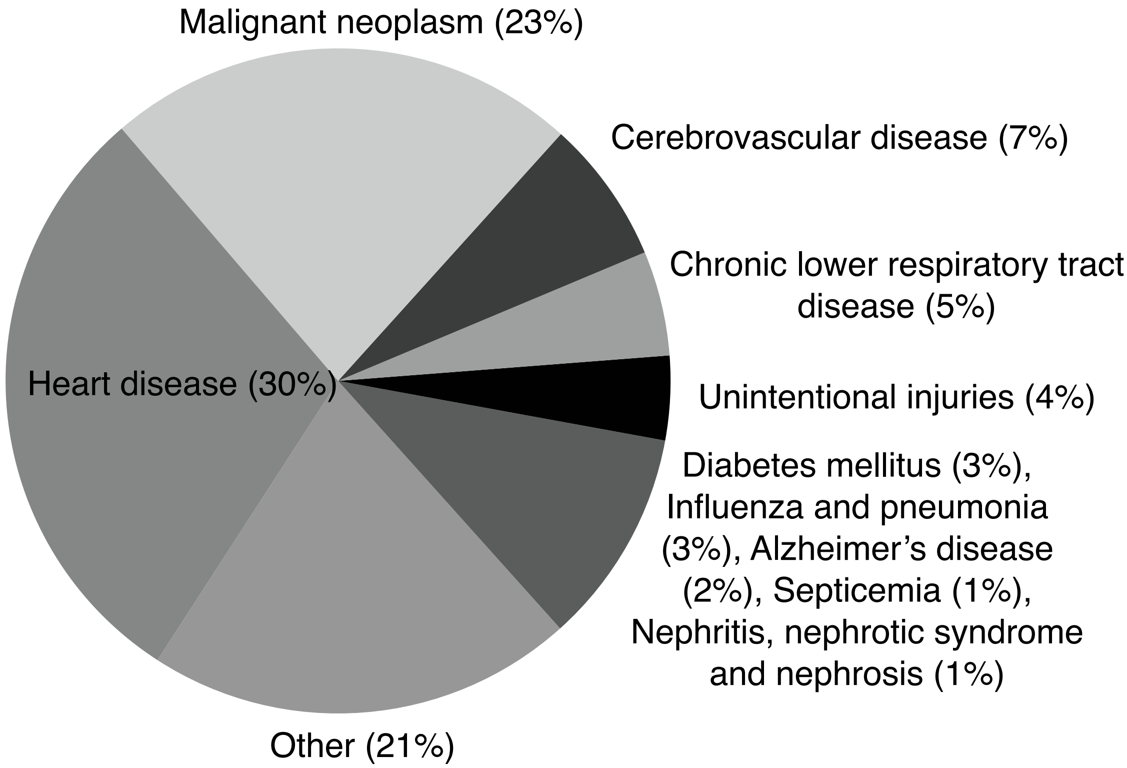
This dissertation is at one intersection between urban planning and public health research. This chapter outlines the motivating theory behind the importance of this intersection. Better planning may not be the sole source of improved health outcomes, but it likely has some role in health behaviors and outcomes given that food provision is in part a spatial behavior requiring people to use and adapt to the built environment.

The Health Problem

The relevant health problem is the basic concern with morbidity and mortality. Because the contributions to death and disease are multi-factorial, parts of this burden may be amenable to urban planning and policy interventions.

Causes of death are often classified by organ system or underlying pathophysiology. In the United States, in the year 2000, the leading causes of death were heart disease (30%), malignant neoplasm ("cancer," 23%), other causes (21%), and cerebrovascular disease ("stroke," 7%) (Mokdad et al. 2004). Thus heart disease, cancer, stroke, along with conditions like type 2 diabetes combined, compose well over half the mortality burden in the U.S. as illustrated in figure 1 below. Although varied in their final disease state, they all have in common the characteristics of being multifactorial in etiology, developing over time, and often being long lasting ("chronic"). Given these origins, they are considered dependent on what are commonly called "lifestyle" risk factors – modifiable individual behaviors like eating, exercise, or smoking.

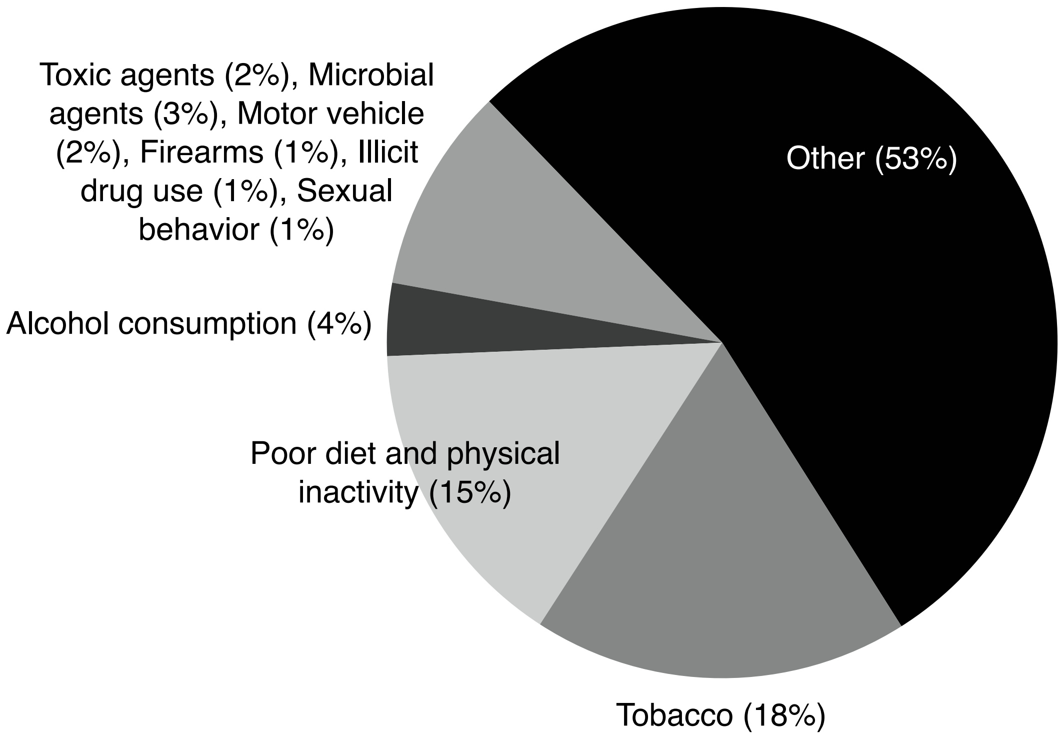
Figure 1: Causes of Death, United States, 2000



Source: Mokdad et al. 2004

Because these underlying risk factors are themselves common, causes of death can be grouped by these risk factors rather than by organ system or pathophysiology. Using this approach, as shown in parallel figure 2 below, the leading "actual" causes of death in the U.S. in the year 2000 were other (53%), tobacco (smoking, 18%), poor diet and physical inactivity (15%), and alcohol consumption (4%) (Mokdad et al. 2004). Within the "other" category, the authors describe a range of probable factors like genetic, biological, environmental pollutants, educational attainment and poverty. Setting the "other" category aside, tobacco and then poor diet and physical inactivity are the leading known causes of death in the U.S.

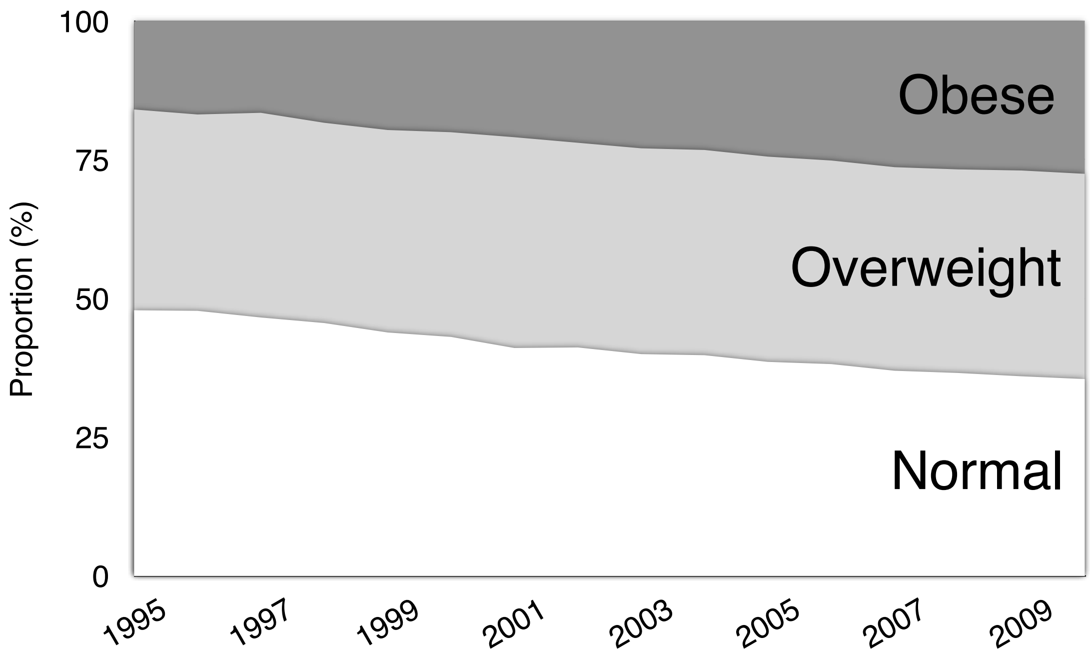
Figure 2: "Actual" Causes of Death, United States, 2000



Source: Mokdad et al. 2004

Poor diet and physical inactivity are considered the primary causes of weight gain, hence weight gain serves as an aggregate proxy for these two risk factors (Mokdad et al. 2004). Body weight is divided into normal, overweight and obese categories based on body mass index. In the U.S. over the past 15 years there have been large increases in the proportion of adults who are classified as obese (Y. Wang and Beydoun 2007). As illustrated in the figure below created from data provided by the Centers for Disease Control and Prevention (CDC) Behavior Risk Factor Surveillance System (BRFSS), the proportion has increased by about 10% over that time, or by about 30 million people. Currently well over half of all adults in the U.S. are classified as overweight or obese (see figure 3 below).[22]

Figure 3: Change in Body Mass Index, United States 1995-2010 (BRFSS)



Source: Behavior Risk Factor Surveillance System (BRFSS, www.cdc.gov/brfss)

This large secular trend in U.S. body mass index does not affect all groups equally. By sex and by race/ethnicity there are consistent differences between

[22] Behavior Risk Factor Surveillance System (BRFSS - www.cdc.gov/brfss)

groups in obesity prevalence (Y. Wang and Beydoun 2007). There are differences between men (see figure 4 below) and women (see figure 5 below) in obesity rates, and both groups have experienced large increases (Y. Wang and Beydoun 2007). By race/ethnicity, there are large differences in women, so that the difference in obesity prevalence for black, Mexican American, and white women is approximately 10% between each group, ranging from about 50% for black women to 30% for white women (Y. Wang and Beydoun 2007). These recent temporal trends, and the differences by sex and race/ethnicity highlight the socially-derived nature of these changes and differences, rather than of genetic or biological etiology.

Figure 4: Male Obesity Prevalence by Race/Ethnicity, United States 1971-2004 (Adapted from Wang 2007)

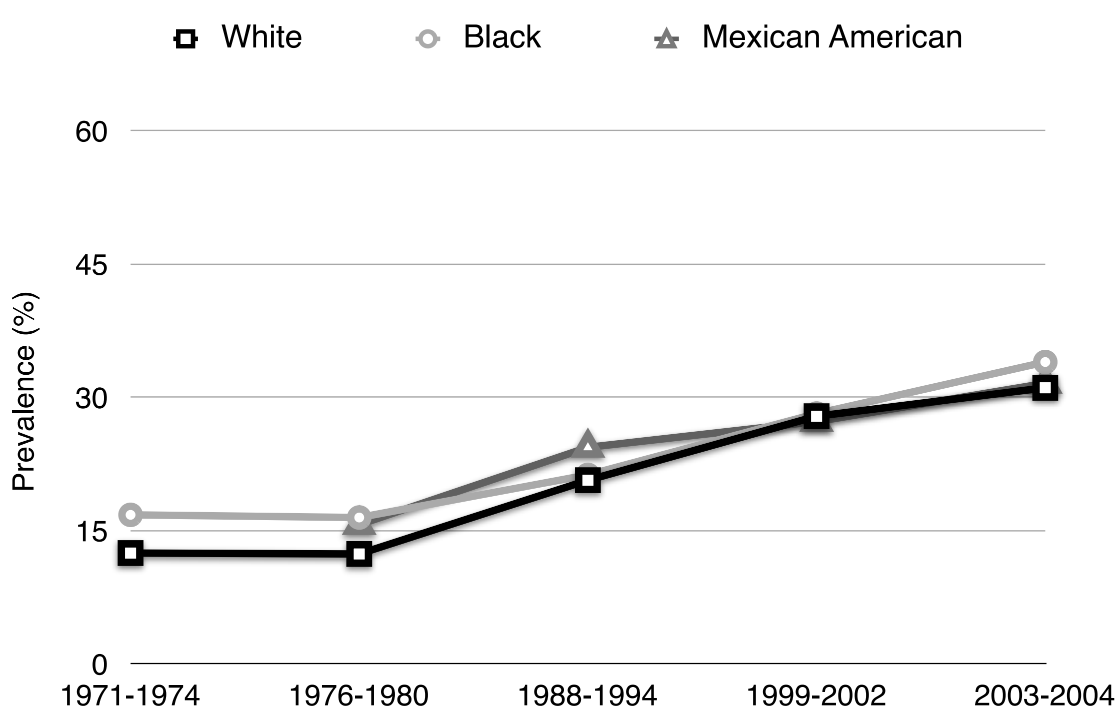
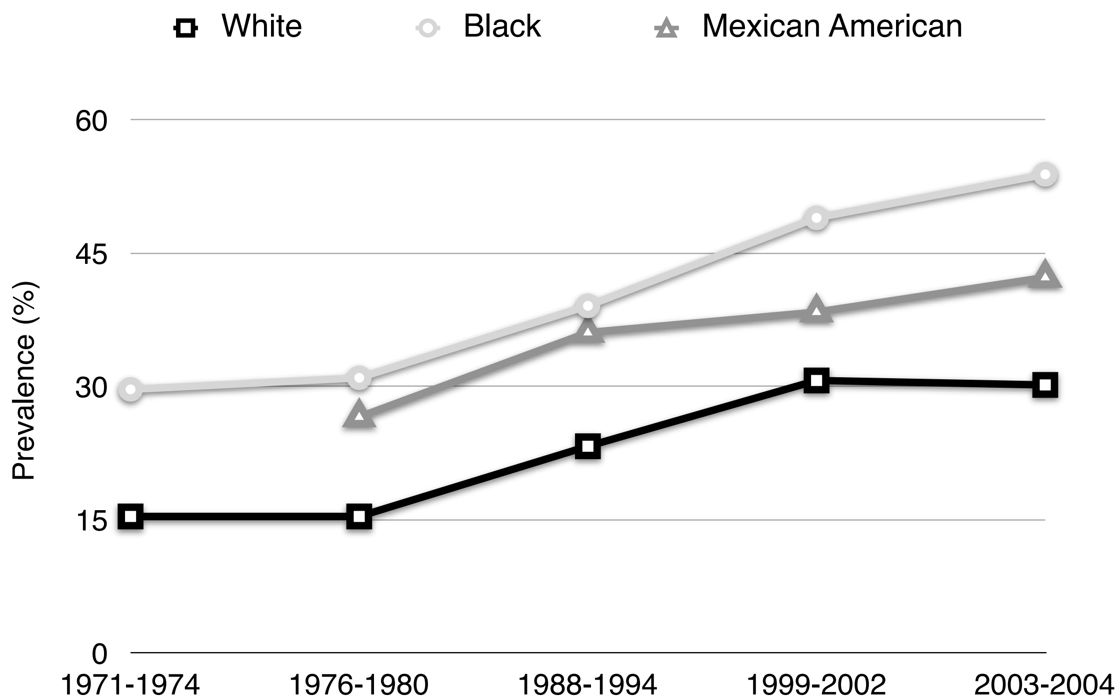


Figure 5: Female Obesity Prevalence by Race/Ethnicity, United States 1971-2004 (Adapted from Wang 2007)



The Planning Problem

The relevant planning problem is the basic concern with whether spatial relationships are an important component to a broad range of outcomes for people and social institutions. Because one of these outcomes may be health, whether there are explicitly spatial health outcomes amenable to urban planning and policy interventions becomes a relevant planning problem.

The urban planning problem can be formulated from two perspectives: 1) the perspective of the professional urban planner employed by a local or regional government authority, and 2) the perspective of the academic urban planner charged with advancing the body of knowledge defined as urban planning within an academy of self-designated members often themselves trained in a diverse range of social sciences.

The professional urban planner faces the challenges of both what to do, and how to achieve those ends. In the history of planning many motivational adjectives have been attached to planning: rational, advocacy, participatory, communicative, and even radical (Friedmann 1987). Despite those threads in planning theory, the professional planner today often develops and implements, or simply mediates the activities of multiple stakeholders in the realm of "spatial public policies and practices" through explicitly rational and communicative means (Huxley and Yiftachel 2000). That process of balancing stakeholder interests in a pluralistic process within bureaucratic or political settings yields planning outcomes. When new stakeholders engage in the process, such as public health or medical care professionals, the opportunity to achieve different outcomes, not possible with the old stakeholders alone, may arise. However, the challenge facing the professional planner is whether these new stakeholders and the newly possible outcomes serve the interests of planning ends however they are defined, be it through rational or other perspectives.

The academic planner can contribute to the body of planning knowledge in several ways. As discussed above, the formulation of planning theory and processes that all planners are taught or engage in is of importance to academic planners. In addition, addressing the question of whether policies, practices or processes have a spatial component is a more basic question facing academic planners. This is important because despite the fact that resources must be distributed in space in some manner - homes, jobs, schools, stores, transportation infrastructure, etc. arranged across the plain we inhabit -- *if* and *how* distribution affects outcomes is the spatial question. If the interaction of people, places and institutions has spatial components, then how those interactions are explicitly spatial should shape the nature of the policies and interventions.

In the end, the questions facing planners are both broad and specific. Is this a spatial problem? Are there components of health problems that are specifically spatial? Does engagement with a new group of stakeholders with new interests advance the multiple ends that define current urban planning practice?

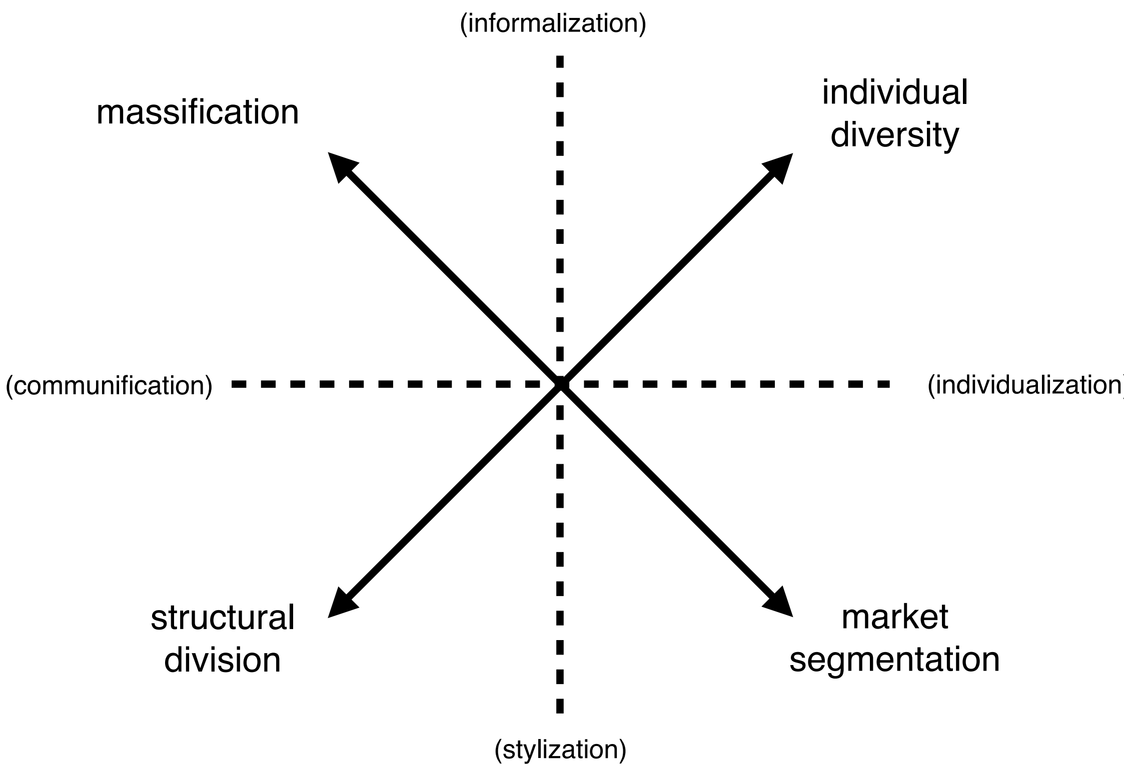
Food Shopping – The Intersection of Health and Planning

Food shopping is one key element at the intersection of health and planning. Eating is a basic requirement to sustain life. Like breathing air and drinking water, without basic nutrition, sustaining life is impossible. However unlike consumption of air or water, eating food has multiple levels of meaning in in personal and social life, so that the act of eating may influence daily life, routines and personal identity. These meaningful factors are often identified in domains of: taste, cost, convenience, health, interpersonal, and larger social factors (Connors et al. 2001; Furst et al. 1996).

Eating as an act of social meaning and influence is well studied by sociologists. Warde (1997) imagines food tastes and consumption varying along two perpendicular axes, one along an axis of individualization to communification and another along an axis of informalization (informal or casual behavior) to stylization. Each quadrant along these axes then forms explanatory theories of changing food consumption or tastes. The four quadrants are massification, individual diversity, market segmentation and structural division. Massification is defined by less individualism (or higher communification) and more informalization. Individual diversity reflects more individualism and more informalization. More individualization and with more stylization defines market segmentation and higher levels of communification and stylization define the structural division quadrant (see figure 6 below).

Massification may be the most familiar explanation for changing tastes. Popular texts like *Fast Food Nation* and popular academics like Marion Nestle argue that changing consumption patterns reflect the "McDonaldization" (Ritzer 1993) of the food system, namely the increased efficiency, calculability, predictability and control of mass produced food. Market segmentation is also another common description of changing food consumption. This describes the evolution of groups who identify with particular patterns of food consumption. The organic food movement can be described this way (Guthman 2003). Structural division theorizes that food consumption changes are due to a reflection of taste based in class structure and division. Food tastes and consumption patterns are reinforced by class divisions and at the same time are in part the forces that maintain those divisions (Bourdieu 1984). Therefore, changing tastes may be the result of class differentiation (Guthman 2003). Alternatively, an explanation built solely on individual diversity suggests that changing food consumption is a reflection of personal expression and knowledge.

Figure 6: "Issues of Taste - Explanations of Changing Consumption" (Adapted from Warde 1997)



In part, to exercise the meaning that eating represents requires individuals to provision food. This food provisioning requires food shopping or patronage at other venues where food can be consumed. Individual resources and knowledge are required for food provisioning, but in addition, the private market creates opportunities for food shopping.

Microeconomic Theory and the Spatial Characteristics of Food Selling

Within microeconomic theory there are four models for the markets that define the relationships between producers and consumers: perfect competition, monopoly, oligopoly, and monopolistic competition. The provision of food is generally considered to abide by the model of monopolistic competition, meaning that there are many producers of a differentiated product with free entry and exit to the market. In this model producers face competition but also have the ability to set prices because competition is not perfect. This is an example of market power (Krugman and Wells 2009).

The ability to achieve market power in a market defined by monopolistic competition is based on product differentiation. There are three main methods of product differentiation: by type, by location, and by quality. Since the food industry is a common example of monopolistic competition, the examples of each form of product differentiation are readily available. Food varies by type, for example based on cuisine, like Chinese versus Italian. These are considered imperfect substitutes, recognizing that for a given individual with preferences one type may be favored over another resulting in a form of market power (Krugman and Wells 2009).

Location is an important differentiator and clearly relevant for this work. Since food must be procured at regular time intervals on a daily basis, the time and costs of travel are incorporated into the costs of food procurement. There are finite limits to the costs that can be incurred to procure food, so that location becomes a *de facto* limitation when selecting food producers. Therefore, from the perspective of the firm, locating in proximity to specific consumers can be a form of market power (Krugman and Wells 2009).[23]

The final type of product differentiation is quality. It is generally accepted that consumers differ in their willingness to pay based on the quality of a product. Thus firms can gain market power by differentiating by the quality of a product. For food, characteristics like freshness, or source (organic, locally produced, etc.) may be differentiators based on quality (Krugman and Wells 2009).

The range of products created by product differentiation is normatively considered of social value (Krugman and Wells 2009). The full range of products created through product differentiation may not be considered of value from specific normative perspectives like health or planning.

Without discussing the theoretical details, because firms in markets characterized by monopolistic competition can charge prices higher than the cost to produce the good, there are incentives for firms to advertise, and for similar reasons, to create brands. In both cases there may be benefits of advertising and branding for the consumer, for example conferring information about a product, but each also represent examples of the market power of firms

---

[23] Location is important for competition among firms at varying spatial scales. For example, Hotelling described the location behavior of firms competing for the same customers. Thus firms make location decisions at multiple scales depending on whether they are differentiating consumers or competing for the same consumers. In the former, location acts as a differentiator, in the latter, location is removed from the differentiating equation.

functioning in markets characterized by monopolistic competition (Krugman and Wells 2009).

Spatial Distribution of People

Almost simultaneously in the mid 1960's several economists began to explore how microeconomic theory could explain urban spatial structure. Notably Alonso (1964) and Muth (1969) developed theories that related housing and transportation costs with distance from the central business district (CBD). These were adapted from nineteenth century agricultural-based theories of location and land rent, with the important advancement to apply utility maximizing theory while accounting for transportation costs to central locations like the CBD. This application assumed transportation costs increased with distance from the CBD, thus housing unit prices had to decrease with distance, or else violate microeconomic theory. This assumption immediately suggests a gradient to housing prices. Also, given assumptions about the income elasticity of housing demand (preference for more housing at higher incomes, backed up by empirical observations), it places higher-income individuals at the periphery and lower-income individuals at the center. Likewise it has implications for changes in both income and transportation costs and the effect that these will have on the distance from the CBD and the quantity of housing purchased. Combined, this theory, gives an explanation for housing price gradients and the position within the city by income groups.

Sociologists have highlighted important institutions and characteristics of markets that can produce unequal spatial outcomes. Massey and Denton (1993) examine the role of racial discrimination (primarily in the housing market) in producing segregation, and the role of segregation in producing concentrated poverty. United Stated Housing and Urban Development (HUD) studies have documented over the past 30 years the clear persistence of racial discrimination in housing markets (U.S. HUD 2000). This discrimination includes firms and agents, as well as lending institutions, local and federal government, zoning and public housing policy. Combined with segregation, economic restructuring also contributed to the creation of concentrated poverty according to Wilson (1987). Both Massey and Denton, and Wilson also take a historical perspective incorporating the effect of major economic and social events (Great Depression, WWII, etc.) on spatial organization. In addition to contemporaneous social conditions, these large social forces and exogenous shocks have long-term effects on spatial organization and differentiation.

Additional economic explanations for spatial differentiation of populations exist. As described earlier, in markets defined by monopolistic competition, location is a key product differentiator, so that in markets for goods that are spatially differentiated, individuals with preferences for those goods may move to be in proximity to them. Within agglomeration theory there is recognition that external economies related to location can be important determinants of clustering for both firms and people (Anas, Arnott, and Small 1998). Tiebout

(1956) also described the effect of multiple municipalities on the provision of public goods. He suggested that resident preferences for a set of public goods could result in a market of municipalities with different public good bundles and hence spatial variation of people. Different levels of transportation resources may also influence the spatial distribution of people (Glaeser, Kahn, and Rappaport 2008). On the other hand, as recognized by Schelling (1971) and more recently introduced to the public health literature by Auchincloss (2011), contexts may represent extreme outcomes unrepresentative of the average level of individual preferences subject to the context because the equilibrium state of the context is dependent on the preferences of a small group.

Travel Mediates Spatial Difference

Food shopping requires travel. Basic "gravity" models[24] have long held power to explain travel behavior and are still in common use among transportation planners. These models describe aggregate travel patterns well using simple aggregate measures of destination size divided by distance to a power factor (Hanson 1995). Specifically describing retail or smaller scale travel, Stouffer (1940) suggested that destination choice could be directly related to the distance to the destination and inversely related to the number of choices between that destination and the individual. Similarly Huff (1963) adapted the gravity model to include the size and variation available at the particular retail

---

[24] Named because it borrows the form describing the gravitational force between two bodies dependent on the mass of each divided the square of the distance between them.

destination. These models imply that food shopping travel behavior is dependent on the aggregate nature of the local built environment, comprised of the number, type and quality of locations and distance to these destinations.

Because transportation resources mediate spatial outcomes, in addition to absolute spatial difference, they can be equally important for explaining disparities in spatial outcomes. Transportation resources may be an important mediator of spatial outcomes both when spatial differentiation is extreme, and when spatial difference is small. Because of this, it is important to explicitly address the independent association of transportation resources on outcomes hypothesized to be spatially dependent. For example in Los Angeles where the majority of the population faces long trips to employment centers, access to private automobiles may be more important for employment outcomes than distance to jobs. Thus it is "transportation mismatch" (Ong and Miller 2005) (differences in transportation resources) which become more important for outcomes than "spatial mismatch" (Kain 1968) (differences in distance, or the spatial distribution of resources).

"Environments"

Modern public health practice embraces the "social ecologic" theory of disease etiology. This theory posits that disease is dependent on social (interpersonal) and "environmental"[25] interactions, which shape the behaviors or

---

[25] The word environment describes interactions commonly conceived of as environmental (from the natural environment) such as air and water pollution, and also metaphorically such as

"lifestyle" of individuals, in addition to biological or genetic origins of disease. Hence, modifying environments that shape behaviors can improve health in addition to, or in lieu of biological interventions, and if implemented among populations may be more effective than individual interventions (Breslow 1996).

Expanding on social ecological theory, Nancy Krieger (2011), at the forefront of public health theory generation, suggests three alternatives to the biological and lifestyle approaches to understanding disease. These are: sociopolitical, psychosocial and ecosocial. Each can be defined independently but also have shared characteristics. Sociopolitical "focuses principally on power, politics, economics, and rights as key societal determinants of health." Psychosocial "emphasizes psychologically-mediated social determinants of population health." Ecosocial "builds on and extends these first two frameworks by analyzing both the embodied population distributions of disease and health and epidemiologic theories of disease distribution, each in relation to their societal, ecological, and historical context" (Krieger 2011, p.163).

The shared characteristics are 1) "the longstanding thesis that distributions of health and disease in human populations cannot be understood apart from — and necessarily occur in — their societal context," 2) "the corollary that social processes causally (albeit probabilistically) determine any health or disease outcome that is socially patterned" 3) "the prediction that as societies

---

interactions with social institutions, or other humanly created social structures, hence the use of quotations. Whether the environmental metaphor is appropriate for relationships formed beyond the natural environment is discussed later.

change, whether in their social, economic, cultural, or technological features, so too will their population levels and distributions of health and disease" (Krieger 2011, p.164).

From Krieger's perspective the most important is the second item, which plainly states that social patterning of disease results from social processes. This is not a "tautology" but rather a critique of the biomedical and lifestyle approach which suggests that social patterning of disease instead originates from behavioral or biological characteristics of individuals (Krieger 2011, p.164).

Conclusions

Urban planners and public health professionals face major challenges as their fields adapt to the changing experience of the people they serve. One example of this is the changing nature of health and disease placed within the larger context of progressive urbanization of the global population.

Because eating is an obligate behavior for life, for most adults, food procurement is as well. Food purchasing is a spatial behavior, in that food stores must be traveled to, and this fact limits the types of foods that might be available for purchase for any given meal. In addition because food is provided by firms in a market characterized by monopolistic competition where location is an important product differentiator and individuals are not evenly distributed in space by characteristics that firms might differentiate upon, food procurement may be further limited to a narrower range of qualities and types because of the market power afforded by location. Thus, in the end, these individual and firm spatial outcomes and interactions may have consequences for health outcomes – the core question addressed by this work.

Chapter 3 – Research Part 1 – Comparison of Reported and Imputed Shopping

Research Question

The first section of this work answers the research question: Does reported chain supermarket shopping behavior match ecological imputation of chain supermarket shopping behavior in Los Angeles County? Reported chain supermarket shopping behavior is defined as the place an individual goes to purchase groceries.[26] Ecologic imputation is defined as the methods used to assign a set of food shopping places[27] as likely places of food purchase based on spatial correlation between an individual's location and locally available food shopping places.

This is an important task for both theoretical and methodological reasons. Theoretically, existing assessments using aggregate measures are subject to ecologic fallacy, or the attribution of causal relationships to lower levels (for example, in individuals interacting with stores) because of observation of association at higher levels (grouping of individuals, or in this case the aggregated shopping potential close to the individual). In addition, it is a methodologically important task because of the effect that measurement error[28] may have on the observed associations between food environments and health outcomes when food environments are imputed.[29] Assessments may be

---

[26] Based on the response from the L.A. FANS survey question.
[27] In this case, chain supermarkets.
[28] This concept is also called errors in variables, or misclassification error.
[29] See appendix B "Measurement Error Models" for a brief discussion.

statistically biased if there is correlation between the imputation method and individual characteristics.

Summary of Findings

The prevalence of supermarkets in home Census tracts is low overall but does vary by level of poverty. Very poor tracts[30] sampled by L.A. FANS have no supermarkets. However despite this outcome, reported supermarket shopping is higher than the prevalence of supermarkets in home Census tracts, with the largest gap between prevalence and reported shopping in non-poor tracts. When supermarkets are present in home Census tracts,[31] the frequency of shopping in supermarkets is higher, however most respondents live in Census tracts without supermarkets, and in this group absence of a supermarket is a poorer predictor of supermarket shopping than the presence of a supermarket in the home Census tract. Comparing multivariate models of imputed to reported shopping, there is evidence that some individual characteristics may bias associations with dependent variables (like body mass index) when only imputed supermarket measures (like tract prevalence) are used.

---

[30] Appendix A contains a description of L.A. FANS poverty definitions.
[31] This choice of geographical unit is discussed later in the chapter.

Literature Review Summary

The goal of this literature review[32] is to summarize ecologic imputation methods described in the literature. It also briefly summarizes the evidence on whether reported shopping behavior matches ecologic imputation and introduces the concept of food place type. Appendix C (Part 1) contains a review of studies.[33] (Block, Scribner, and DeSalvo 2004; Powell, Slater, et al. 2007; Horowitz et al. 2004; Moore and Diez Roux 2006; Zenk et al. 2005; Morland et al. 2002; Morland, Diez Roux, and Wing 2006)

Ecologic imputation requires two decisions, 1) assignment of spatial attributes to the individual and assignment of spatial attributes to the food places, 2) a decision about the degree of spatial correlation relevant for the underlying process for which imputation is being undertaken.

Individuals are commonly assigned their residential location as the spatial attribute. This can be the exact residential location (home address) or the approximated residential location when exact address is unknown or populations are the primary interest. Approximated residential location is commonly a Census tract or tract centroid.

---

[32] The National Cancer Institute publishes an online database called "Measures of the Food Environment" (https://riskfactor.cancer.gov/mfe/), which contains a systemic review of all English-language publications from 1990 to present that attempt to measure the food environment (using multiple approaches). The databases currently contains (as of Sept 1, 2012) 598 articles. To prioritize this list of studies for this review, I linked the studies indexed by the National Library of Medicine (via Pubmed ID, or PMID) to a database of citation counts.

[33] Appendix C lists all food environment specific studies discussed in this proposal. The Part 1 review lists the top seven studies in the NCI database. The first study, by Morland has been cited 511 times, more than twice the next frequently cited article. As can be seen in the summary, each demonstrates a unique approach to defining food environments.

Food places are assigned their spatial attributes, either based on exact location or approximated by either Census tract or ZIP code membership.

Spatial correlation can be measured in several ways. The most common method is by Census tract. In that case ecologic imputation of food places is determined by common membership to a Census tract, of individuals (or populations) and food places. In this case, the count of food places, or the density of food places creates the measure, with either tract population or tract area as the denominator. Alternatively, ZIP codes can be used in a similar manner.

Alternatives to predetermined zones (Census tracts or ZIP codes) are radii. In this approach, a circular buffer of radius x is used to define an area for assignment of spatial correlation to the individual. Individual location may be determined exactly or approximated by a residential zone centroid. Using this approach, exact food place locations are typically used. Counts or densities are then calculated to determine the ecological measure. Radii used in the literature include distances of 0.5 mi, 1 mi, 1km, 3 km, 5 km and 8 km.

Separate from common zone assignment (either radii or tracts/ZIPs) distances between the individual and food places can be measured. For this approach the individual assignment is a point in space, either the exact location or zone centroid, and the food place is the exact location. Distance is measured via the street network or in a straight line to the nearest food place (or to the nearest x places).

Each of the above measures can be created based on food place types. Examples of food place types are supermarkets, grocery stores, convenience stores, fast food restaurants, full service restaurants, etc. Ratios of the measures can also been created.

In the seven studies cited as examples in the literature of studies using imputed ecologic measures, there are 6 different approaches to creating the imputed measures. Studies use different geographies, Census tracts, ZIP codes, and locally-defined neighborhoods. Within these geographies stores are assigned to these geographies either by count, population normalized counts, or distance to the centroid of the geography. Thus the literature clearly demonstrates the conceptual challenge outlined in table 1 below.

Table 1: Matrix of Approaches to Food Place Ecologic Imputation Measure Creation

		Individual	Populations	
Food Place Relative Spatial Relationship		Residential (exact)	Residential (zone)	Other***
Tract	Count (Y/N)	*1 (for t)**	**	*1 (for w, t)**
	Count (n)	*2*	**	*2*
	Density (pop)	*3*	**	*3*
	Density (area)	*4*	**	*4*
ZIP	Count (Y/N)	*5*	**	*5*
	Count (n)	*6*	**	*6*
	Density (pop)	*7*	**	*7*
	Density (area)	*8*	**	*8*
Radii	Count (Y/N)	*9 (for x, t)**	*15*	*9 (for w, x, t)**
	Count (n)	*10*	*16*	*10*
	Density (pop)	*11*	*17*	*11*
	Density (area)	*12*	*18*	*12*
Distance	Nearest	*13 (for t)**	*19*	*13 (for w, t)**
	Nearest n	*14*	*20*	*14*

The intent of the table is to enumerate the number of approaches available to the researcher to create imputed measures based on a rational decision about meaningful spatial importance and the available information on individuals or populations. Each approach is numbered sequentially 1-20 for residential location as the origin, or 1-14 for other locations as the origin.

* Each measure can be divided by food place types (t) (supermarkets, fast food, etc.), divided into multiple radii of distance (x) (where appropriate), and/or be based at an alternative origin (w), creating additional dimensions to the table.

** Yields the same results as exact residential assignment if spatial association is determined by common membership, otherwise the measure falls into the "other" column.

*** Non-residential location like work (w), school, or activity space.

Some studies have noted the characteristics of reported shopping behavior. Across several studies in the public health literature, (Chaix et al. 2012; Drewnowski et al. 2012; Inagami et al. 2006) urban planning literature, (Clifton 2004; Handy 1996; Hillier et al. 2011) and in government reports, (Mantovani and Welsh 1996; Ohls et al. 1999) these studies conclude that individuals generally shop both inside and outside of their neighborhood (or Census tract) and often beyond the nearest store.

The results from the review of ecologic imputation methods and the studies which contain insight on reported shopping behavior, lend empirical support to the theoretical argument that comparing ecological imputation to reported food shopping behavior is relevant to understanding the association between food environments and health. There is no current example in the literature of a study that directly compares reported shopping behavior to imputed measures of shopping behavior.

Aim of Comparative Analysis (in lieu of a Hypothesis)

The main aim of part 1 is: To compare the magnitude of consistency between reported food shopping outcomes and the ecologic imputation of this outcome. This includes examining pattern of inconsistencies overall and within population sub-groups. Similar to analysis of diagnostic medical testing, false positive and false negative rates will be compared. It is not feasible to test all imputation methods outlined in the table above, so this proposal with focus on the first method which is use of Census tracts.

Data

I use three data sources for the part 1 analysis. The United States Census (year 1990) is the source of geographical zones (Census tracts).

Population data from the 2000 Census is approximated for 1990 geography. The InfoUSA business database (year 2003) is the source for food places. It contains listings of food places classified by North American Industrial Classification System (NAICS) with street addresses and exact geocodes (latitude/longitude). The Los Angeles Family and Neighborhood Survey (L.A. FANS) is the source of reported shopping behavior. In 2001-2002, L.A. FANS asked approximately 2,600 households where they shopped for groceries, recording the store name and location (closest major street intersection).

Methods

The goal of the methodological approach is to compare ecologic imputation to reported behavior. Individual and local covariates are introduced to help explain correlations.

The matrix above in table 1 summarizes the approaches to ecological imputation performed in the literature to date. For a single store type, there are 20 different ecologic imputation approaches which could be created for a given exact individual residential location. This analytical section tests just the first case, Census tract, because it is the most common geography used in the literature compared to the other geographies that could be tested.[34]

---

[34] The literature review summary table contained in the appendices summarizes the geographical unit used in each study.

I subdivided food places by type[35] as done previously in the literature. The primary subtype of interest is supermarket defined by North American Industrial Classification System (NAICS) code and sales volume (>$2M sales). This is further subdivided by chain status. Chains are defined by having more than 10 locations in the region and corporate ownership.[36]

The primary aim of ecological imputation is to estimate whether an individual shops at a chain supermarket. The most basic construction of the observed outcome, compatible with that imputation, is whether an individual shopped in a chain supermarket, irrespective of the actual location of that supermarket.

The imputed outcome is shopping in a chain supermarket. There may be many ecologic states (in a given geography) that could result in imputing shopping in a supermarket (or not shopping in a supermarket):

1) If there is no chain supermarket, but an alternative food place type (for example smaller grocery store) in the geography then shopping is imputed not to be a chain supermarket. If the ecologically imputed probability is given by P^i, then $P^i = 0$.

---

35 The concept of type simply refers to a grouping process to aggregate store brands or locations. Thus types can originate from industrial classifications, store characteristics like sales volume, or simply grouping of stores by the same name. In this research, types take different forms based on each of these approaches and depending on the analytical aim. Each approach is described in detail later in each method section.

36 There is no formal definition of a supermarket "chain." I describe the detailed enumeration of supermarket chains later in this section.

2) If there is one chain supermarket and no alternative food place in the geography, then shopping is imputed to be chain supermarket. If the ecologically imputed probability is given by P^i, then $P^i = 1$.
3) If there is one chain supermarket and other alternative food places in the geography, then shopping can be imputed by two methods:
   a) If the ecologically imputed probability is given by P^i, then $P^i = 1$, or
   b) a continuous probability.[37]
4) If there are many chain supermarkets and other alternative food places in the geography, then shopping can be imputed by two methods:
   a) If the ecologically imputed probability is given by P^i, then $P^i = 1$, or
   b) a continuous probability.[38]

If imputation method 3(a) and 4(a) are used then a 2x2 table can be constructed (table 2 and table 3) that compares the counts of observed shopping and ecologically imputed shopping. For example:

[37] See Appendix D "Size and Distance-based Store Probabilities"
[38] See Appendix D "Size and Distance-based Store Probabilities"

Table 2: Matrix of Possible Outcomes From Observed Shopping Behavior Compared to Imputed Shopping Behavior

	Observed (1 = shopped in chain supermarket)	Observed (0 = did not shop in chain supermarket)
Imputed (1 = results from method 2, 3, 4 above, shopped in chain supermarket)	a = count of individuals observed to shop in chain and imputed to shop in chain supermarket (“true positive”)	b = count of individuals observed not to shop in chain supermarket but imputed to shop in chain supermarket (“false positive”)
Imputed (0 = result from methods 1 above, did not shop in chain supermarket)	c = count of individuals observed to shop in chain supermarket but imputed not to shop in chain supermarket (“false negative”)	d = count of individuals observed not to shop in chain supermarket and imputed not to shop in supermarket (“true negative”)

Table 3: Matrix of Possible Outcomes From Observed Shopping Behavior Compared to Imputed Shopping Behavior With Hypothetical Results

	Observed (1 = shopped in chain supermarket	Observed (0 = did not shop in chain supermarket)	Total
Imputed (1 = results from method 2, 3, 4 above, shopped in chain supermarket)	1400	100	1500
Imputed (0 = result from methods 1 above, did not shop in chain supermarket)	300	200	500
Total	1700	300	2000

Part 1 contains three approaches to examining consistency of the observed and imputed chain supermarket shopping:

1) Directly modeling the correct test assignment as a function of neighborhood and individual characteristics,

2) Directly modeling observed chain supermarket shopping as a function imputed chain shopping, individual and neighborhood characteristics, a precursor to a full measurement error modeling approach using model results from Part 2,
3) Implementing a diagnostic medical testing approach comparing the sensitivity and specificity of the test overall and within population subgroups.

Multivariate regression models test associations between neighborhood and individual characteristics and the ability of the ecologically imputed measure to correctly match observed behavior. The dependent variable in the models is the difference in the probability (P) between the observed (P^o) and the imputed (P^i), given by ($P^o - P^i$). For the case when chain supermarket shopping is imputed to 1 or 0, this methods results in a dependent variable which identifies the correct assignment (either chain supermarket shopping, or not, cell a or cell d in the table 2 above) as a dichotomous outcome.[39] These are "true positives" and "true negatives."

I estimate multivariate regression models with observed chain supermarket shopping as the dependent variable. This is the first step in a measurement error modeling approach, often called a validation study. In this

---

[39] When the probability is assigned based on retail sales, a continuous measure ranging from 0 to 1 results. See Appendix D "Size and Distance-based Store Probabilities."

model the ecologically imputed shopping measure is included along with individual and neighborhood characteristics used in Part 2 of this work.

Several independent variables can be introduced to the models at the neighborhood level (Census tract) and for individuals (described in table 4).

Table 4: Theoretical and Operationalized Measures Used in the Modeling Approach to Assessment of Ecologic Measure Validity

Measure Class	Conceptual Measure	Operationalized Measure	Comment
Dependent Variable	Predictor of shopping in a place that is health promoting or detrimental to health	Concordance between observed and imputed chain supermarket shopping [dichotomous]	Some evidence to support correlation between store type and health (i.e. chain supermarkets are health promoting). Store type is based on a single report of store name and location as the most frequent location of grocery shopping.
Neighborhood characteristics	May be independently associated with presence of a chain supermarket	Population density, racial/ethnic composition, % car ownership, poverty [categorical]	
Individual characteristics	May be independently associated with selection of a chain supermarket	Household income [continuous], Household car ownership [dichotomous]	
Other factors	Effect of Imputation Error	Distance from residential location to centroid (if using area for imputation)	

* The functional forms of measures used in the final analysis are noted in brackets.

Model Example:

Food place type (supermarket, non-supermarket) = imputed food place type + distance to store + individual covariates + neighborhood covariates + survey structure covariates, given by

$$y_{ij} = \beta_0 + \beta_n x_{nij} + \beta_m x_{mj} + (u_{0j} + e_{ij})$$

where x_{nij} is a vector of n individual level predictors and x_{mj} is a vector of m tract level predictors. If the dependent variable is a categorical outcome then the model will be in the logit form.

Based on the 2x2 contingency table presented above, standard measures of diagnostics test performance are calculated. These measures are sensitivity, specificity, positive predictive value, and negative predictive value. The sensitivity and specificity (or 1 - specificity) can be plotted for a given method of testing (or sub-group) for comparison. This generates an "ROC" (receiver operating characteristic) space with points. If a continuous classifier is used, for example distance, then a full ROC curve is generated.

Sensitivity is defined as the effectiveness of a test to correctly identify individuals with the outcome (in this case shopping at a chain supermarket). It is calculated (using the labels from table 2) as: Sensitivity = a / a + c or TP / TP + FN. Specificity refers to how effective the test is at identifying individuals without the outcome. It is calculated as: Specificity = b / b + d or FP / FP + TN. There are overall standards for test performance, and as stated above, comparing these metrics between tests allows for evaluation of relative test performance. (Loong 2003)

Store Type Creation

To define chain supermarkets, I acquired the InfoUSA business directory for year 2003 and reviewed all stores within the NAICS 445110x[40] classification for Los Angeles County using a web-based application I developed. I grouped identical store names and calculated frequency, average annual sales and employee counts. I assumed stores with the identical name to be from the same store chain. The primary criteria for chain selection is location frequency in Los Angeles County. There is no formal chain supermarket definition, but chain membership has been defined by both national corporate ownership and frequency of location.[41] Store groupings based on identical names were grouped further based on similarity of names and by considering sales and employee counts. Typically chain supermarkets are defined by sales of greater than $2M with employee counts in the range of 30 to 50 employees.[42] If a store had a

---

[40] The description for NAICS 455110 is "Supermarkets and other grocery (except convenience) stores." The "x" indicates subgroups within this classification.

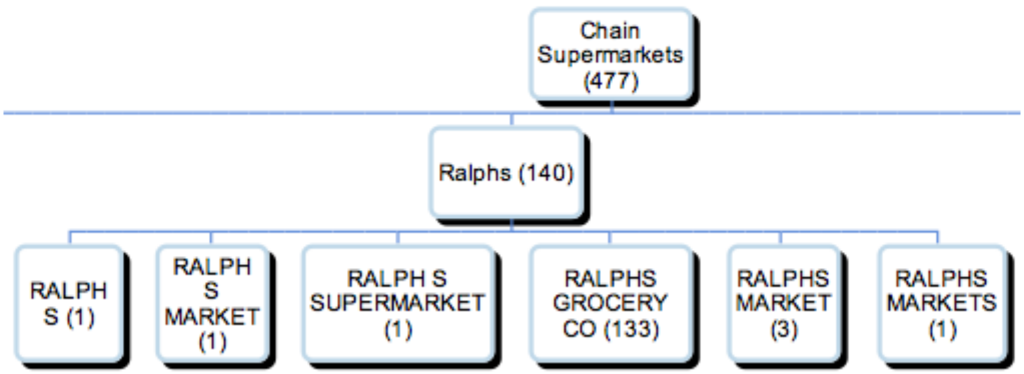
[41] For example in Morland (AJPM 2002) chains are defined as "large, corporate-owned" and in the industry publication *Progressive Grocer* its annual industry report divides sales by store count with the group comprising 11 or more locations titled "chain." (http://www.progressivegrocer.com/inprint/article/id2694/glass-half-empty-/, accessed May 23, 2013)

[42] The $2 million annual sales cut off is a common supermarket definition as identified by industry groups (http://www.fmi.org/research-resources/supermarket-facts, accessed May 23, 2013), and the United States Department of Agriculture, for example in the recent report "The Extent of Trafficking in the Supplemental Nutrition Assistance Program: 2006-2011" (http://www.fns.usda.gov/ora/MENU/Published/snap/ProgramIntegrity.htm, accessed May 23, 2013)

similar name to a chain supermarket and annual sales and/or employee count in these ranges in was included in the chain supermarket definition.

Each chain supermarket name is a common name developed by the author based on grouping of similar store names. For example the Ralphs chain supermarket is defined by several synonyms represented in figure 7 below.

Figure 7: Ralphs Chain Supermarket and Synonyms in InfoUSA Database



Apostrophes have been removed from the names to aid matching. The majority of locations are named ‘RALPHS GROCERY CO’ but other names such as ‘RALPHS MARKET’ are also used. Other examples can be viewed using the interactive application. Synonyms were only included if annual sales or employee counts were similar to the primary chain name of high frequency.

Chain supermarkets were geocoded to 1990 Census tracts using geocodes provided by InfoUSA. Store locations lacking a geocode, or with a geocode flag indicating a poor match as provided by InfoUSA were re-geocoded using the Google Maps geocoding application programming interface (API).[43]

Detailed geographic boundaries are not readily available for 1990 Census tracts. There are two approaches to geocoding, a database and address matching approach and a coordinate point (after address geocode) in geometry approach. While the former may be possible by recreating address and tract assignments from original Census data, this was not possible given limited resources. Instead, I obtained detailed (versus cartographic) Census boundaries based on original Census data from the National Historical Geographic Information System (NHGIS) at the University of Minnesota. I then transformed them from their original projection to the Keyhole Markup (KML) format (WGS 84 projection) adopted by Google Maps as a file format. I did this in GRASS[44] an open source Geographic Information System (GIS) package. Once in KML format, I placed the file in a Google Fusion Table. I assigned Census tracts to store points defined by latitude and longitude (with a 1 meter buffer as required by the application) using the Google Fusion Table API ST_INTERSECT command. When more than one tract was returned by the query, the first result returned was used for tract assignment.

L.A. FANS sampled from 65 1990-based Census tracts. With chain supermarkets defined and geocoded to tracts, the L.A. FANS tracts with chain supermarkets can be defined. This is the definition of imputed chain shopping --

---

[43] The majority of supermarkets used the InfoUSA geocode and less than 10% used the Google Maps geocode

[44] The Geographic Resources Analysis Support System; Available at http://grass.osgeo.org

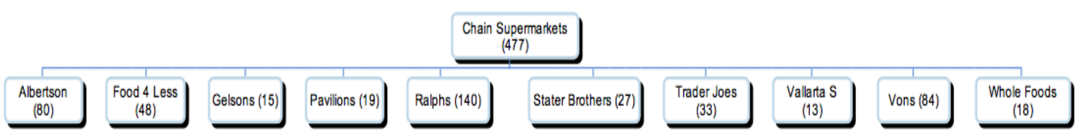
if a tract contained a chain supermarket all residents in the tract are imputed to shop at a chain supermarket.

Results

The aim of Part 1 is to define chain supermarkets, create a method of imputing chain shopping status, and compare imputed chain shopping status to reported chain shopping status in the L.A. FANS wave 1 cohort.

The final chain supermarkets defined for Los Angeles County were Albertsons, Food 4 Less, Gelsons, Pavilions, Ralphs, Stater Brothers, Trader Joes, Vallarta, Vons, and Whole Foods for a total count of 477 chain supermarkets. Figure 8 below displays the results of the chain supermarket grouping process.

Figure 8: Final Chain Supermarket Definition

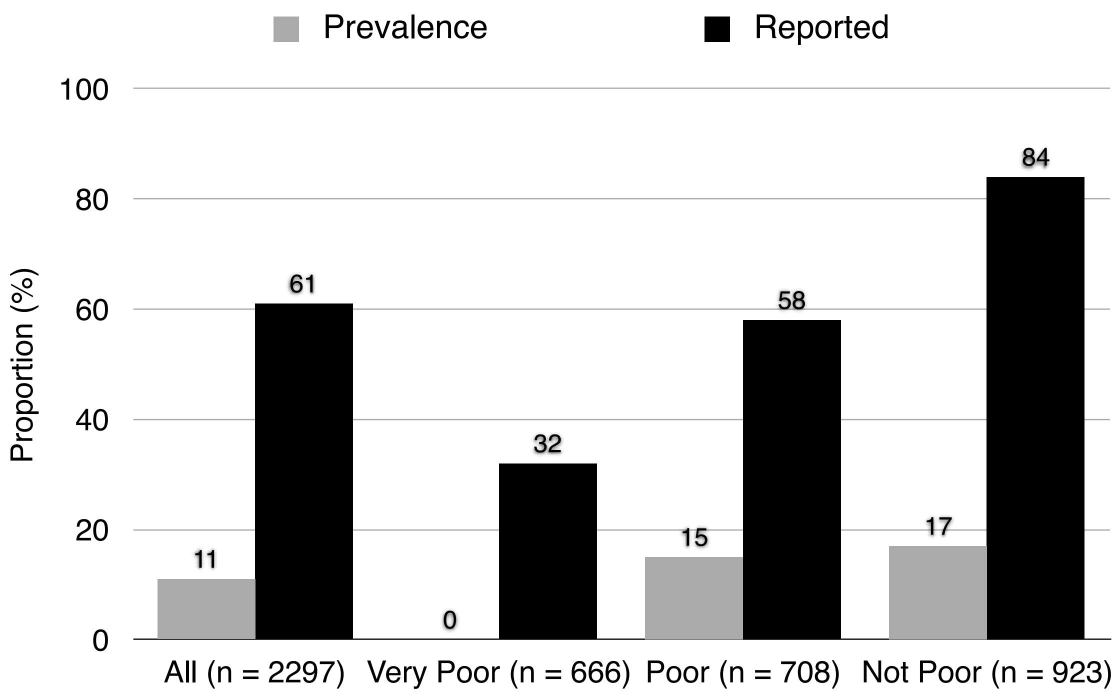


This figure along with other candidates excluded from selection can be viewed using an interactive application developed by the author.

Based on this approach, 7 tracts of the 65 contained chain supermarkets. L.A. FANS sampled 20 very poor, 20 poor and 25 non-poor tracts. Of the 20 very poor tracts, no tracts contained chain supermarkets, of the 20 poor tracts, 3 contained a chain supermarket, and of the 25 non-poor tracts, 4 contained a chain supermarket

Figure 9 below shows the tract prevalence and reported frequency of chain supermarket shopping among respondents.

Figure 9: Chain Supermarket Shopping Rate (Tract Prevalence versus Reported)



In the full sample 11% of respondents are to be expected to shop in a chain supermarket based on the prevalence of chain supermarkets in home Census tracts, but instead 61% report shopping in chain supermarkets. When this is divided by the L.A. FANS poverty categories, in the very poor tracts, no tracts contained chain stores so no respondents are imputed to shop in chain supermarkets. However, 32% of respondents shop in chain supermarkets. In the poor tract category imputed and reported shopping rates are similar to the total sample overall, at 15% and 58% respectively. The non-poor tract group has an imputed chain shopping rate of 17%, but a reported chain shopping rate of approximately 84%.

Sensitivity and specificity are alternative approaches to compare the observed result of a test (in this case imputation of chain shopping) to the gold standard for that test (reported chain shopping). The results show that imputation of chain shopping by census tract is a highly specific test but not sensitive. This indicates that if the test is positive, there is fairly high confidence that an individual shops in the chain supermarket. However the low sensitivity indicates that when the test is negative, it is unclear whether an individual may shop in a chain supermarket. Detailed results are presented in table 5.

Table 5: Sensitivity and Specificity of Imputation in Main Sample and by L.A. FANS Poverty Category

All Results	Observed(1)	Observed(0)	
Expected(1)	238	24	262
Expected(0)	1162	873	2035
	1400	897	2297
Sensitivity (TP / TP + FN)	0.17		
Specificity (TN / TN + FP)	0.97		
PPV (TP / TP + FP)	0.91		
NPV (TN / TN + FN)	0.43		
Very Poor	Observed(1)	Observed(0)	
Expected(1)	0	0	0
Expected(0)	211	455	666
	211	455	666
Sensitivity (TP / TP + FN)	0		
Specificity (TN / TN + FP)	1		
PPV (TP / TP + FP)			
NPV (TN / TN + FN)	0.68		
Poor	Observed(1)	Observed(0)	
Expected(1)	92	17	109
Expected(0)	318	281	599
	410	298	708
Sensitivity (TP / TP + FN)	0.22		
Specificity (TN / TN + FP)	0.94		
PPV (TP / TP + FP)	0.84		
NPV (TN / TN + FN)	0.47		
Not Poor	Observed(1)	Observed(0)	
Expected(1)	146	7	153
Expected(0)	633	137	770
	779	144	923
Sensitivity (TP / TP + FN)	0.19		
Specificity (TN / TN + FP)	0.95		
PPV (TP / TP + FP)	0.95		
NPV (TN / TN + FN)	0.18		

PPV = Positive Predictive Value, NPV = Negative Predictive Value, TP = true positive, TN = true negative, FP = false positive, FN = false negative

A series of multivariate models assess the degree of association between reported shopping behavior, imputed shopping behavior, individual and neighborhood characteristics of the L.A. FANS sample.[45] The detailed results are discussed here and summarized at the end of this section.[46]

The first series of analyses compares models with reported chain shopping as the dependent variable. These models assess the association between individual, household and neighborhood characteristics and reported shopping behavior. The analysis compares models selected empirically using backward stepwise selection and full models with all covariates conceived to be theoretically important for the dependent variable used in main analysis from part 2 (body mass index). In all models except for the final model in this set of analyses, imputed chain shopping is a covariate.

As the results indicate, in both parsimonious and full models, there are several predictors that are statistically significant. These are: Latino ethnicity (compared to all other groups, and compared to white race alone), less than high school educational attainment (compared to all other groups and compared to a college degree or higher attainment), US born status, and the poverty category of the Census tract, divided into very poor, poor and non-poor (the referent category

---

[45] The inclusion of individual and neighborhood characteristics are based upon the characteristics selected for the main analysis in part 2 and their selection are detailed there. The independent variables in this section are not conceived to have any *a priori* association with the dependent variable. In part the aim is to identify any associations between the independent variables used in part 2 and reported chain shopping.

[46] Detailed tables with results for each model are in Appendix F.

is non-poor). All covariates are negatively associated with reported chain shopping except for US born which is positively associated. Chain shopping imputation is positively associated with reported chain shopping. Any direction of association is important for identifying biased predictors, as discussed later.

The next analysis also compares the results of models with observed chain shopping as the dependent variable. In this case, the analysis compares full models with and without chain shopping imputation as the primary independent variable, as in the previous analysis, except this time comparing L.A. FANS survey Census tract poverty categories to a continuous Census tract disadvantage score. The results indicate that less than high school educational attainment (compared to college of higher) is negatively associated, and U.S. nativity is positively associated with observed chain shopping, with or without control for imputed chain shopping (presence of a chain store in the home Census tract). Latino ethnicity is negatively associated or unassociated with observed chain shopping depending on the measure of poverty or neighborhood disadvantage used as control. In models with neighborhood disadvantage as a covariate, less than high school educational attainment (compared to college or higher) is negatively associated with observed chain shopping only when imputed chain shopping is included in the model.

The next analysis is similar to the previous analysis, except that it includes imputed chain shopping as a dependent variable to compare with models containing observed chain shopping as the dependent variable. Because imputed chain shopping is perfectly correlated with L.A. FANS Census tract poverty categories, only a continuous neighborhood disadvantage score can be used as a control variable. For the L.A. FANS sample, imputed chain shopping (presence of a chain supermarket in the home Census tract) was negatively associated with the age of the individual. Depending on the control variables used, other individual and neighborhood characteristics are associated with observed chain shopping as discussed previously.

The final analysis in this series compares the results of models with matches to imputed chain shopping defined by both cases of correct assignment - correctly determining chain shopping or correctly determining absence of chain shopping. The results compare parsimonious models theoretically conceived based on the L.A. FANS survey structure to full models determined by theoretical associations with chain supermarket shopping. In the full models, less than high school education and high school educational attainment (compared with college or higher education) are positively associated with correct imputation. Household income is also associated with correct imputation. The continuous measure of neighborhood disadvantage is negatively associated with correct assignment, but the Census tract poverty categories are associated in either direction with correct assignment.

The summary table below compares full models with each dependent variable, observed chain shopping, imputed chain shopping, and matched imputation (yes and no). Neighborhood disadvantage is used to account for

Census tract level poverty because of the perfect correlation between imputed chain shopping and L.A. FANS Census tract poverty categories. The results of each model have been described above. Depending on the dependent variable, different individual and neighborhood characteristics are associated with the outcome of interest. The pseudo-R-squared values for each model with reported shopping is in the range of 0.2, relatively low. However the pseudo-R-squared values for models using imputed chain shopping are an order of magnitude lower.

Table 6: Comparison of Observed, Imputed, and Matched Chain Shopping Controlling for Individual and Neighborhood Characteristics

Chain Shopping Reported	Dep. Var.		
Chain Shopping Imputed		Dep. Var.	
Chain Shopping Match yes-no			Dep. Var.
Model Number	2012-12-6-10	2012-12-6-11	2013-1-17-10
Observations	2297	2297	2297
Parameter Count	15	15	15
Wald Chi Squared	212.19	58.22	90.59
Chi Squared Test P-Value	0	0	0
Log Likelihood	-1061.5421	-1168.0782	-1383.6989
Pseudo R-squared	0.2079	0.0285	0.144
Age (years)	1.00 (0.99-1.01)	0.99 (0.98-1.00)*	1.00 (0.99-1.01)
Female	1.06 (0.71-1.59)	0.99 (0.79-1.26)	0.88 (0.65-1.17)
Household with Children	0.93 (0.64-1.35)	1.17 (0.74-1.84)	1.31 (0.87-1.99)
Latino	0.54 (0.28-1.02)	0.66 (0.26-1.62)	1.05 (0.53-2.07)
African American or Black	0.80 (0.36-1.79)	0.67 (0.20-2.29)	0.78 (0.34-1.76)
Other Race or Ethnic Group	1.10 (0.58-2.07)	1.13 (0.44-2.95)	0.87 (0.43-1.75)
Education - Less Than High School	0.58 (0.29-1.17)	1.86 (0.70-4.90)	2.91 (1.53-5.53)*
Education - High School	0.90 (0.44-1.86)	2.25 (0.95-5.30)	1.96 (1.00-3.84)*
Education - Some College	1.13 (0.61-2.09)	1.47 (0.58-3.68)	1.27 (0.67-2.39)
Household Income (dollars)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)*
Employed	0.99 (0.68-1.43)	1.09 (0.77-1.54)	1.17 (0.82-1.67)
Married or Living with Partner	1.02 (0.76-1.37)	1.14 (0.80-1.60)	0.95 (0.67-1.34)
US Born	2.46 (1.78-3.40)*	1.37 (0.69-2.75)	0.73 (0.45-1.19)
Household Owns Automobile	0.76 (0.52-1.13)	1.22 (0.87-1.70)	1.25 (0.90-1.74)
Neighborhood Disadvantage (Wink.-Cub.)	2.19 (1.42-3.38)*	0.83 (0.34-2.00)	0.54 (0.30-0.96)*
[Constant]	4.64 (1.56-13.75)*	0.17 (0.03-0.85)*	0.48 (0.16-1.50)

* $p < 0.05$
Results are odds ratios and 95% confidence intervals. All models are survey weighted and account for clustering by census tract. Race and ethnicity categories are Latino, white, African

American or black, and other category is primarily composed of Asian groups. Education categories are less than high school, high school, some college, and college or higher. Poverty categories are very poor, poor, and non-poor. When interpreting parameter estimates the referent category is the category not listed in the table within the same column and may be more than one category. In tables with multiple models the referent categories may differ between models.

Conclusions

The primary contribution of this section is to define a method for testing the association between ecologically imputed measures and reported behavior. This has not been systematically conducted in the literature to date. This systematic approach also allows for assessment of ecologically imputed measures and conclusions about whether their use is appropriate when considered against reported behavior.

In this large sample of Los Angeles adults drawing from a broad range of neighborhood poverty strata are found insights about the use of imputed supermarket measures compared to reported supermarket measures and the potential limitations of using imputation alone. In the 65 sampled 1990 Census tracts, only 7 contained a major chain supermarket as commonly defined based on data provided by InfoUSA. There are clear differences in the poverty strata in the prevalence of chain supermarkets – there are none in very poor tracts, but 32% of respondents in very poor tracts report shopping in supermarkets. Conversely the rate of reported supermarket shopping in non-poor respondents is 84% despite only 17% of this group residing in a Census tract with a chain supermarket. These extreme examples demonstrate the limitation of this approach and the broad misclassification in both directions of assignment that may take place with the use of imputation.

Conceived as a medical test or screening tool, the knowledge that an individual lives in a tract with a chain supermarket may be associated with shopping in a chain supermarkets. However, the majority of the sample lives in tracts without chain supermarkets, and in this group, the test is of limited utility.

The measurement error modeling approach also contributes to understanding the limitation of imputation. This modeling approach, with reported chain shopping as the dependent variable and imputed chain shopping as the main independent variable of interest helps identify estimators, which may be biased. When other independent variables are associated with statistical significance to the dependent variable, reported chain shopping, while controlling for imputed chain shopping, models using imputed chain shopping as independent variable to predict another outcome (for example body mass index) and including the statistically significant predictors can be assumed to be biased by these predictors. Because there are several predictors that are statistically significant, these results indicate that models with dependent variables such as body mass index should be cautious when including both imputed chain shopping (presence of a chain store in the Census tract) along with these common individual and neighborhood characteristics.

Chapter 4 – Research Part 2 – Body Mass index and Store Type

Research Question

This section of the analysis answers the research question: Is reported shopping behavior in specific food store types associated with body mass index (BMI), *ceteris paribus*? Reported shopping behavior is defined as the place an individual purchases food. Food store types are classifications of stores along administrative measures, industrial classifications (supermarkets, grocery stores, fast food, etc.), or other characteristics. The health outcome, body mass index, is a ratio of weight to height.

Summary of Findings

In multivariate models with body mass index as the dependent variable and store types as the main independent variable, specialty and Spanish-language store types are associated with statistically significant lower body mass index, controlling for individual and household characteristics as well as the sampling Census tract poverty strata.

Literature Review Summary

The goal of this literature review[47] is to: 1) Establish the association between specific food store types (for example, supermarkets) and health

---

[47] The National Cancer Institute publishes an online database called "Measures of the Food Environment" (https://riskfactor.cancer.gov/mfe/), which contains a systemic review of all English-language publications from 1990 to present that attempt to measure the food environment (using multiple approaches). The databases currently contains (as of Sept 1, 2012) 598 articles. To

outcomes (diet, body mass index), 2) Highlight that in most cases this association is based on ecologic imputation of food store type.[48]

Most studies use individually measured body mass index as the outcome, not population averages.[49] All studies subdivide food stores by type. For example, supermarkets, grocery stores, fast food restaurants, full service restaurants, and convenience stores are the most common categories. All but two of the studies (Chaix et al. 2012; Drewnowski et al. 2012) use an imputed ecologic measure of food stores (as discussed in part 1) divided by store types. Nine of the 31 studies focus on adolescents or children. Several were longitudinal, in that they examined change in the BMI outcome in relation to change in the ecologically imputed measure. Most studies adjust for individual and neighborhood covariates. Two studies included transportation resources as a covariate. The tables in Appendix C (Part 2 literature review) summarize the results.[50] (Jeffery et al. 2006; Inagami et al. 2006; M. C. Wang et al. 2007; Lopez 2007; Powell, Auld, et al. 2007; Brown et al. 2008; Millstein et al. 2009; Powell and Bao 2009; Rundle et al. 2009; Galvez et al. 2009; Inagami et al. 2009; Rose et al. 2009; Zick et al. 2009; Black et al. 2010; Jilcott et al. 2010; Laska et al.

---

prioritize this list of studies, I linked the studies indexed by the National Library of Medicine (via Pubmed ID, or PMID) to a database of citation counts.

[48] Appendix C contains a complete summary of the literature review.

[49] For example, population average BMI in a neighborhood or county.

[50] This part of the literature review was based on a search for body mass index (or BMI) in each abstract. This identified 73 studies. Reviewing each, 31 studies with relevant outcomes for this part of the proposal were identified. The review focused on studies including supermarkets as a food place, studies conducted in the United States, and select studies focusing only on fast food when other aspects of the study were relevant for this part of the proposal. Studies are presented in the Appendix C table sorted by citation count.

2010; Ford and Dzewaltowski 2011; Gregson 2011; Hickson et al. 2011; Casagrande et al. 2011; Dubowitz et al. 2011; Jilcott et al. 2011; Chaix et al. 2012; Keegan et al. 2012; Wall et al. 2012)

Of the 31 studies, I classified 13 as having generally positive outcomes. This means that the food store measure (subdivided by type) is associated with BMI in the expected direction of association, for example more supermarkets per area is associated with lower BMI. I classified another 13 studies as having mixed results, in that the associations observed between BMI and the food place measures were in the expected direction of association for only some food place types or for only some sub-populations. I classified the remaining 5 studies as negative studies meaning they found no association between BMI and ecologically imputed measures, or that the direction of association was in the opposite direction of expected association.[51]

With regard to food store type, only a single study (Chaix et al. 2012) creates sub-types within supermarkets (except for the distinction between chain and non-chain supermarkets found in some studies). A recent study subdivided stores into three categories based on market basket price (Drewnowski et al. 2012). Otherwise, for all other studies, the primary hypothesis tested is whether supermarkets are associated with lower BMI.

---

[51] This is not a representative sample of all studies in the database so definitive conclusions about the distribution of the most common direction of association cannot be made.

Hypothesis

The main hypotheses for part 2 are: Observed shopping behavior in a food store classified as a chain supermarket is associated with lower body mass index after adjusting for individual and neighborhood characteristics. Observed shopping behavior in a food store classified as a discount supermarket is associated with higher body mass index after adjusting for individual and neighborhood characteristics. No *a priori* hypothesis is made about the association between body mass index other food store types.

Data

I use three data sources for the part 2 analysis. The United States Census (year 1990) is the source of geographical zones (Census tracts) and local (neighborhood) covariates (year 1997 estimates). The InfoUSA business database (year 2009) is the source for food places. The Los Angeles Family and Neighborhood Survey (L.A. FANS) is the source of observed shopping behavior, individual covariates and the outcome of interest, body mass index. In 2001-2002, L.A. FANS asked approximately 2,600 households where they shopped for groceries, recording the store name and location (closest major street intersection), covariates and the self-reported outcome (height and weight).[52] Additionally in 2007-2008, L.A. FANS followed up with the same grocery store

[52] Self-reported BMI may be another source of measurement error.

question among approximately 1,200 households, added several diet and physical activity related questions, and directly measured body mass index.

Methods

The goal of the methodological approach is to describe the association between body mass index and reported food store type. Body mass index is calculated based on measured height and weight performed by the L.A. FANS research staff during wave 2 respondent interviews.

Food store location and type is based on the store name reported in L.A. FANS and the location. Store name and location are linked to the InfoUSA 2009 database to verify locations and to gather additional information for store typing, specifically annual sales volume.

Food store types are created based on NAICS industrial classification, store names, and sales volume. Based on NAICS code and sales volume (>$2M sales) a broad supermarket category is created. This is the first classification of food store type, supermarket versus other food store type. Store brands with 10 or more stores of the same name present in Los Angeles County created the chain supermarket category.

From the chain supermarket category, a subgroup of “major” English-language chains is grouped to create the major chain category. Grouping food store names with “less,” “save,” or “bargain” in the name, created the discount food store type. Stores with a Spanish-language word in the name are grouped into a category called Spanish-language chains. Stores with fewer locations, smaller format, specific product focus, and/or limited product inventory formed a specialty store category. Stores outside of the above categories but larger in sales than smaller local markets (>$2M annual sales) formed the independent store category. Large regional stores selling groceries and often other products formed the bulk category. Small stores with lower annual sales (<$2M annual sales) formed the small market category. If a reported store was not located in the InfoUSA database but a similar store (by name) was assigned a type then it was assigned the same type. Because there is evidence that major chains vary in their store contents by the income, poverty status, or racial/ethnic composition of the store's local area, (Sloane et al. 2003; Horowitz et al. 2004) the major chain category was stratified by level of Census tract poverty.

I calculated distance to food store along the street network using ArcGIS 10 and the current version (provided with ArcGIS 10) of North American street network grid as provided by ESRI. This distance is included as a covariate in models.

Individual covariates include: age, sex, race/ethnicity (Latino, black, white, other), U.S. born, Spanish-language interview, education (less than high school, high school, some college, college), employment, family income (with imputation status), married/partner, and household car ownership. Health-related covariates include: ever exercising in past week, eating fast food in prior day, and current smoking status. The theoretical justification for inclusion of these covariates is described in table 7 below.

Other covariates in the model include elements of the survey sampling scheme, which are poverty strata (very poor, poor, not poor), and whether the household contained children. A neighborhood disadvantage index consisting of educational attainment, family income, housing value, occupational status, and employment status was tested in some models as a substitute for neighborhood poverty (Winkleby and Cubbin 2003).[53]

Table 7: Theoretical and Operationalized Measures Used in the Modeling Approach to Body Mass Index

Measure Class	Conceptual Measure	Operationalized Measure	Comment
Dependent Variable	Health	Body mass index (weight/height ratio, kg/m^2) [continuous]	This measure is self-reported. It is possible there may be systematic differences in reporting among some groups. Of particular concern would be if reporting was biased by the independent variable of interest
Independent variable of interest	Typology of stores, or store attributes, that are either health promoting or detrimental to health	Store typed by industrial classification, size, and other characteristics hypothesized to be associated by health [categorical]	Some evidence to support correlation between store type and health. Store type is based on a single report of store name and location as the most frequent location of grocery shopping.

53 The variable was substituted for the L.A. FANS provided poverty strata because L.A. FANS poverty estimates were based on Los Angeles County Urban Research 1997 poverty estimates created from 1990 Census data and local population change adjustments, whereas the constructed variable uses 2000 Census data a temporally closer estimate to the sampled time period.

Measure Class	Conceptual Measure	Operationalized Measure	Comment
Control variables – individuals	Mutable and immutable characteristics that may be independently associated with the dependent variable and if unaccounted for mask associations with the independent variable of interest	Age, Sex, Educational attainment Race/Ethnicity, Employment, Household Income, Nativity	These are derived from the prior literature, which has examined the associations between body mass index and individual characteristics.
Control variables – neighborhood	Mutable and immutable area level characteristics that may be independently associated with the dependent variable and if unaccounted for mask associations with the independent variable of interest	An index of educational attainment, family income, housing value, occupational status, and employment status (Winkleby 2003).	These are derived from the prior literature, which has examined the associations between body mass index and area level characteristics.
Survey Structure	The survey oversampled specific populations which must be accounting for in the statistical analysis	Poverty Strata (Very Poor, Poor, Non-Poor) [categorical]; Households with children (Y/N) [dichotomous]; Survey weight	

Measure Class	Conceptual Measure	Operationalized Measure	Comment
Other variables – individual	Characteristics which may be associated with the independent variable of interest	Household automobile availability [dichotomous], Household size [continuous], Distance to store [continuous]	
Other variables - neighborhood	Characteristics which may be associated with the independent variable of interest	Population density, local competition (distance between two closet large stores), proportion going to a specific type of store in the sample	

* The functional forms of measures used in the final analysis are noted in brackets.

I estimated survey weighted multilevel multivariate linear regression models (MLMs). MLMs are used to account for the correlation among observations (in this case sampling by census tract), to introduce covariates at that level, and partition variance between levels.[54]

Model Example:

Model 1 – Body Mass Index = food place type (supermarket, non-supermarket) + distance to store + individual covariates + neighborhood covariates + survey structure covariates, given by

[54] See Appendix F for additional detail.

$$y_{ij} = \beta_0 + \beta_n x_{nij} + \beta_m x_{mj} + (u_{0j} + e_{ij})$$

where x_{nij} is a vector of n individual level predictors and x_{mj} is a vector of m tract level predictors.

In addition to MLMs, I estimated ordinary least squared (OLS) regression models with survey weights and adjustment of standard errors for clustering of observations by Census tract.[55]

Results

This section presents the results from Part 2 analysis, which includes descriptive statistics from the analysis sample, model estimates, and several sensitivity analyses.

<u>Descriptive Statistics</u>

Table 8 below gives examples of stores included in the store type classification scheme.

Table 8: Store Name Examples by Store Type

Store Type	*Example*
Discount	Food 4 Less, Payless Foods
Specialty	Trader Joe's, Whole Foods
Spanish-Language	Superior, Vallarta
Major	Albertsons, Ralphs, Vons
Independent	Jons
Bulk	Costco
Small	Jerry's Market*

[55] The results from OLS models are presented because of the limitations in partitioning survey weights between levels in MLM estimates. Additionally, the store type coefficients are the focus of this analysis so inference about the variance is not required, commonly the intent of MLM estimates.

*Hypothetical name to protect respondent confidentiality

The main analysis sample consisted of 915 participants from L.A. FANS wave 2 respondents. The sample represents 35% of the original 2619 households sampled in wave 1. The mean body mass index in the sample is 26.9 kg/m^2. The median distance traveled to reported store is 1.14 miles. Overall the sample is 58% Latino, 25% white, 9% black or African American, and 8% other racial/ethnic groups, comprised primarily of Asian respondents. These are summarized in table 9.

Table 9: Sample Characteristics - BMI, Age, Sex, Race/Ethnicity

	L.A. FANS Wave 2
N (%)	915 (35%)
BMI - kg/m^2, mean (SD)	26.9 (5.2)
Distance to store - miles, MD (IQR)	1.4 (0.57-1.9)
Age - years, mean (SD)	39.7 (12.8)
Female	61%
Race/Ethnicity	
Latino	58%
White	25%
Black	9%
Other	8%

Educational attainment is divided into four categories. The most frequently reported response is less than high school education, in 35% of the sample. Table 10 below summarizes other characteristics of the sample including family income, employment, nativity, marital status, smoking status and household car ownership.

Table 10: Sample Characteristics - Education, Income, Employment

	L.A. FANS Wave 2
Educational Attainment	
Less Than HS	35%
High School	18%
Some College	25%
College or Higher	21%
Family Income -$1,000 median (IQR)	30 (15-58)
Employed	68%
Smoker	13%
U.S. Born	46%
Married/Partner	65%
Own Car	70%

<u>Store Type Shopping Frequency</u>

Store responses are divided into type categories as described previously. Major chain stores are the most frequently reported store type, in 37% of respondents. I divided the major chain category in half according to rank by 1997 poverty estimates in 1990 Census tract geography. This division at the median for the entire group creates two groups stratified by poverty and yielded a median percent in poverty of 5.7% in the lower poverty major chain stores and 14.6% in the higher poverty major chain stores.

The second most frequently reported store type is the Spanish-language chain, at 27% followed by discount stores at 17%. Specialty stores are shopped in by 5% of the analysis sample. Independent, bulk stores and small markets are shopped in at similar proportions. Median poverty of the store tract is 7.7% in specialty stores, slightly higher than in major chain stores. Median tract poverty in Spanish-language chains is 28%. Table 11 below summarizes store type frequency and median poverty level of the Census tracts in which stores were located for select store types.

Table 11: Store Type Frequency and Median Tract Poverty

	L.A. FANS Wave 2	Store tract median poverty
Discount	152 (17%)	
Specialty	48 (5%)	7.7 %
Spanish Language	243 (27%)	28%
Major	333 (37%)	
Major (high SES)	160 (18%)	5.7%
Major (low SES)	173 (19%)	14.6%
Independent	40 (4%)	
Bulk	63 (7%)	
Small	33 (4%)	

Descriptive Statistics Stratified by Race/Ethnicity

Later analyses by racial/ethnic stratification are reported so descriptive statistics are included for background. The mean BMI for Latino respondents was 30.2, and for black respondents 30.1, compared to 27.7 for white respondents and 26 for other respondents. There are large differences by race/ethnicity in the proportion of respondents in each of the tract poverty sampling strata. For example 54% of black respondents live in very poor tracts and 83% of white respondents live in non-poor tracts. Among Latino respondents, 38% live in very poor tracts and 41% live in poor tracts. Similarly the median household income differs by approximately $60,000 between white households and black or Latino households. Among Latinos, 24% report being U.S. born and 55% attained less than a high school education. Table 12 below summarizes the sample characteristics stratified by race/ethnicity.

Table 12: Sample Characteristics - Stratification by Race/Ethnicity (1)

	Latino	White	Black	Other
N	533	227	84	71
BMI - kg/m2	30.2 (6.0)	27.7 (5.5)	30.1 (7.3)	26.0 (6.7)
Neighborhood Poverty				
Very Poor	38%	1%	54%	8%
Poor	41%	16%	14%	21%
Non Poor	21%	83%	32%	70%
Income - $1000 median (IQR)	35 (20-58)	95 (42-167)	36 (23-83)	67 (28-113)
Age - years mean (SD)	43.8 (12)	50.8 (14)	47.0 (13)	50.6 (14)
US born	24%	86%	94%	31%
Less than HS	55%	7%	13%	1%

Frequency of store type shopping varies by race/ethnicity. Among Latino respondents 42% shop in Spanish-language store chains, 22% shop in discount named stores, and 20% shop in major chain stores. Among black respondents, 50% shop in major chains (36% in major chains in higher poverty Census tracts), 21% in discount named stores and 18% in Spanish-language chains. Among white respondents, 55% shop in major chains (39% in major chains in lower poverty Census tracts) and 17% in specialty chains. Among other respondents 55% shop in major chains, as well as 13% in bulk stores and 14% in smaller markets. Table 13 below summarizes the store type frequency by race/ethnicity.

Table 13: Store Type Frequency - Stratification by Race/Ethnicity

	Latino	White	Black	Other
Discount	22%	5%	21%	6%
Specialty	1%	17%	2%	6%
Spanish Lang.	42%	1%	18%	4%
Major	20%	55%	50%	55%
high SES	7%	39%	14%	30%
low SES	13%	26%	36%	25%
Independent	5%	3%	2%	3%
Bulk	6%	7%	6%	13%
Small	4%	2%	0%	14%

Distance traveled to the reported stores is similar across racial/ethnic groups. White and black respondents travel a similar median distance, 1.6 miles. Latino respondents travel 1.3 miles to the store and other respondents 1.4 miles. Car ownership varied across racial/ethnic groups. In Latino households, 60% report owning a car, compared to white, other and black households, which owned cars at a 75%, 72%, and 70% rates, respectively.

Table 14: Distance to Store and Household Car Ownership by Race/Ethnicity

	Latino	White	Black	Other
Distance to store - miles median (IQR)	1.3 (0.7-2.3)	1.6 (0.8-2.7)	1.6 (0.8-2.6)	1.4 (0.9-2.4)
Own Car	60%	75%	70%	72%

<u>Multivariate Model Results</u>

I estimate multivariate models with body mass index as a function of individual, household, and neighborhood characteristics. I introduce store types to estimate any additional association with BMI these type categories may contribute in addition to individual, household and neighborhood characteristics.

Full models control for: age, sex, race/ethnicity (Latino, black, white, other), U.S. born, Spanish-language interview, education (less than high school, high school, some college, college), employment, family income (logged, imputation status), married/partner, household car ownership, distance to store over street network, ever exercising in past week and eating fast food in prior day, smoking status, and survey sampling strata, the tract poverty level (very poor, poor, and non-poor) and households with children.

The table below shows selected parameter estimates for fully adjusted models with and without store type.[56] Among individual and household characteristics smoking status is associated with an approximately two point lower BMI. Family size is associated with a 0.53 point increase in BMI, so that for each additional family member BMI increases 0.53 points. For respondents in very poor tracts, BMI is 2.5 points higher in models with store type control, compared to respondents in non-poor tracts. The BMI of respondents in poor tracts is 1.7 points higher in models with store type control, compared to respondents in non-poor tracts.

Adding store type to the full model results in statistically significant differences in some store types compared to the referent category, major chain stores in higher poverty tracts. The store types included in the model are: major chain (higher and lower poverty), discount named store, Spanish-Language chain, specialty chains, independent supermarkets, bulk stores, and small stores.

[56] Full results with all parameter estimates are available in Appendix G.

For respondents shopping in specialty chains BMI is 2.8 points lower compared to shoppers in major chains in higher poverty tracts. For respondents shopping in Spanish-language chains BMI is 2 points lower compared to shoppers in major chain stores in higher poverty tracts. A global F test of the store type categorical variable was not statistically significant.[57] Table 15 below summarizes the multivariate model results comparing the full model with and without store type adjustment.

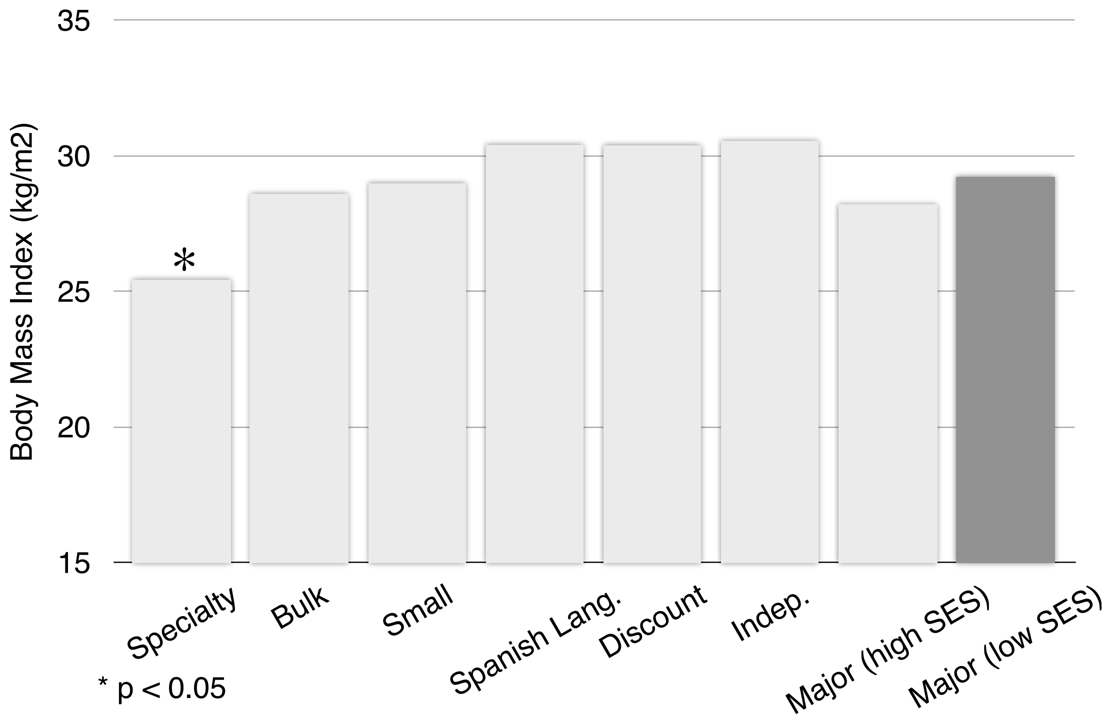
Table 15: Multivariate Model Results - Body Mass Index and Store Type

	Beta	p-value	Beta	p-value
Individual/Household Characteristics				
Smoke	-2.0	0.015	-1.9	0.008
Family Size	0.53	0.025	0.53	0.025
Tract Characteristics (ref: non-poor)				
Very Poor	2.2	0.023	2.5	0.011
Poor	1.5	0.030	1.7	0.031
Store Type (ref: major in higher poverty tract)				
Specialty	-	-	-2.8	0.014
Spanish Lang.	-	-	-2.0	0.029

Figure 10 below depicts the average BMI of shoppers within each store type unadjusted by any covariates and figure 11 below depicts the average BMI of shoppers within each store type after adjustment for all covariates in the full model.

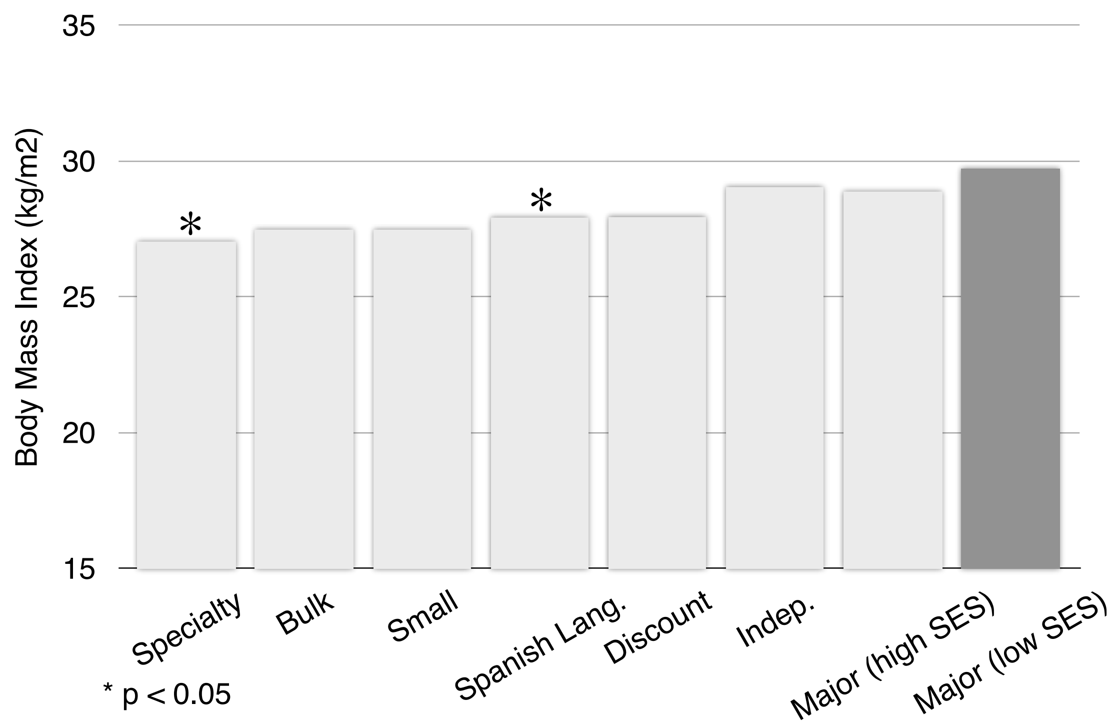
[57] The identical test in a larger sample based on self-reported BMI from Wave 1 respondents was statistically significant. Sensitivity analysis are discussed later in the section and full results presented in Appendix G.

Figure 10: Unadjusted Body Mass Index by Store Type



Mean shopper body mass index (unadjusted) is similar in all categories except for the specialty chain in which BMI is significantly lower than the referent category, major chains in higher poverty tracts.

Figure 11: Adjusted Body Mass Index by Store Type

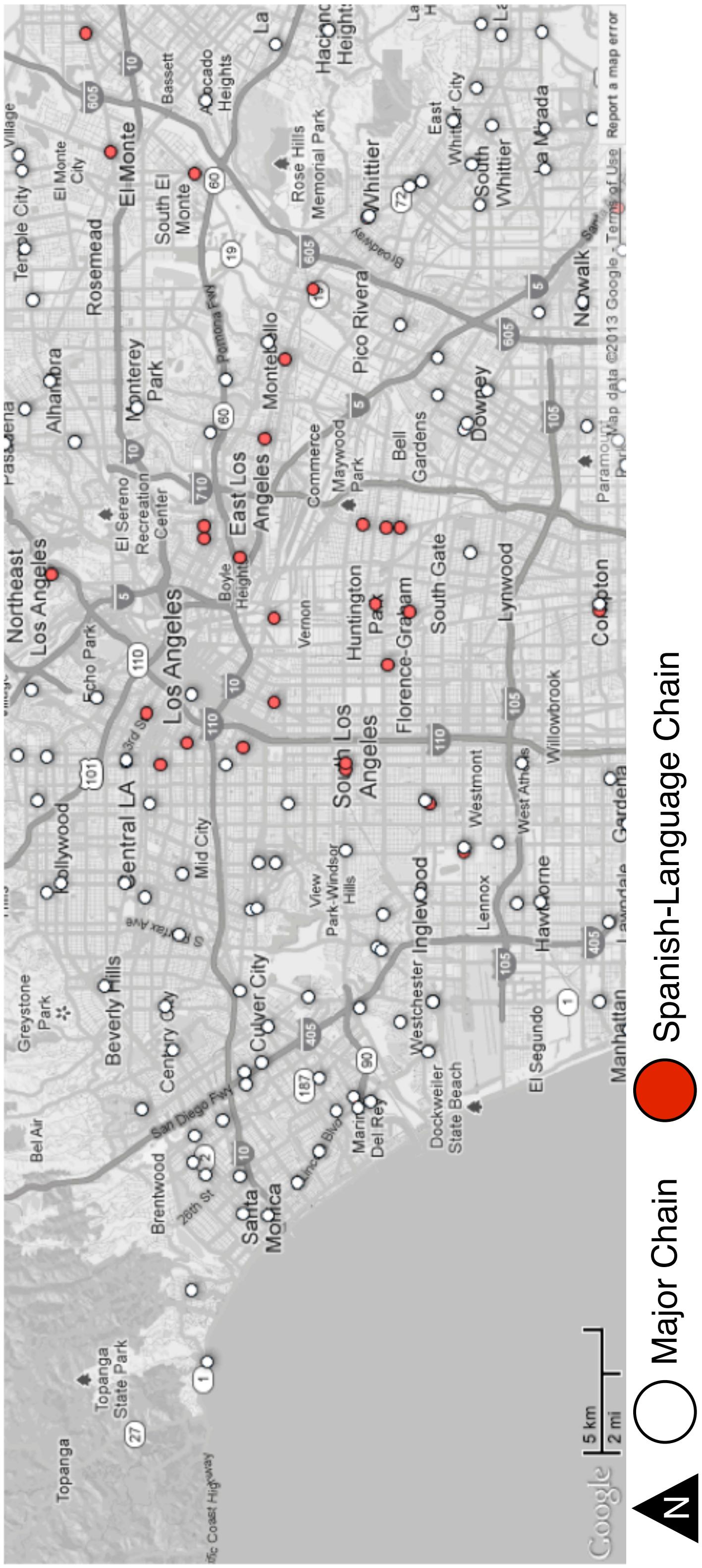


As shown in Figure 11, specialty store and Spanish-language chain shoppers are associated with lower average BMI compared to shoppers in major chains in higher poverty tracts.

Store Type Maps

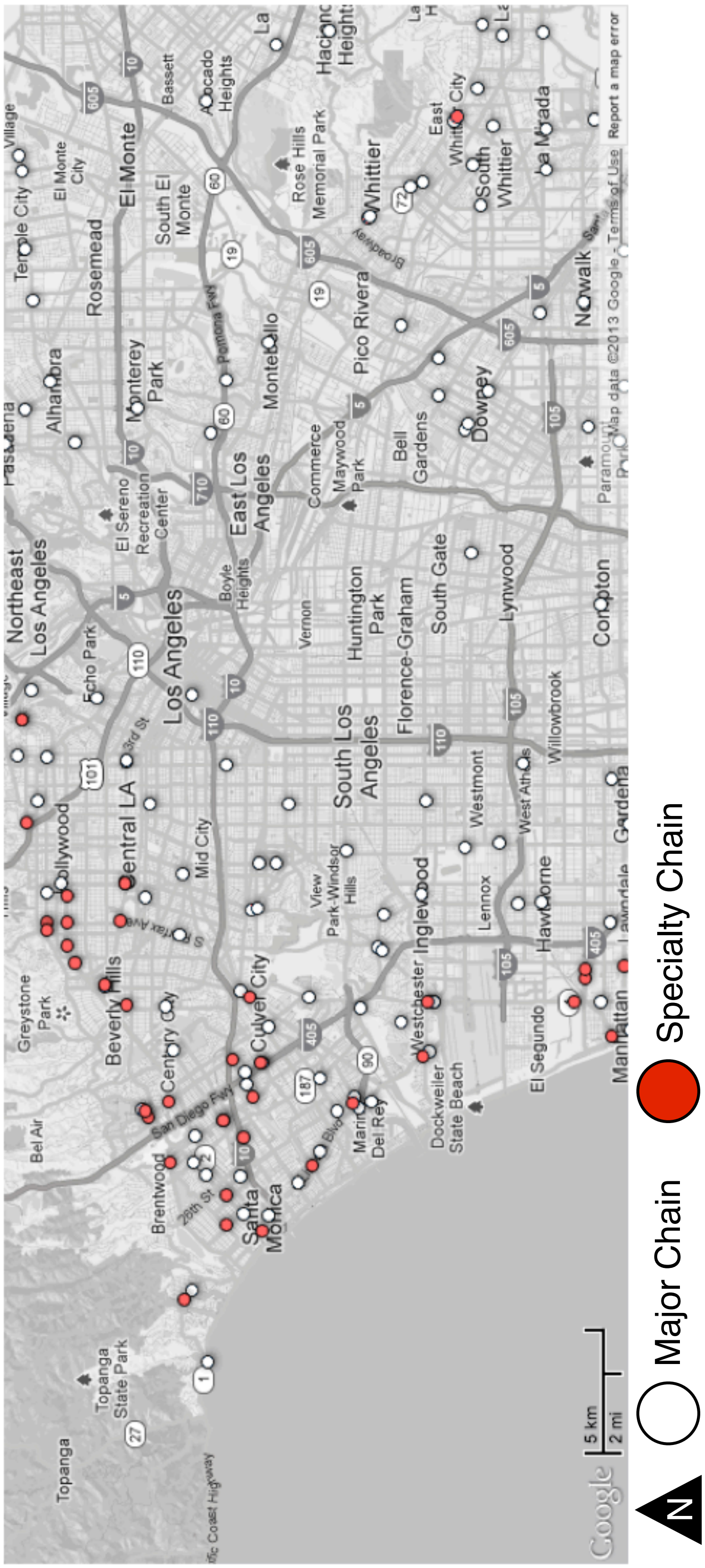
The following maps (figure 12 and 13 below) compare the locations of store types with statistically significant differences in average shopper body mass index. In the first map (figure 12), points in red (dark gray in print) show Spanish-Language chains compared to major chains in white. Locations are based upon all stores in the InfoUSA 2009 database and not all stores depicted are included in L.A. FANS responses. The Spanish-Language chains are geographically clustered in a region including Huntington Park and East Los Angeles. In parts of South Los Angeles and west of downtown Los Angeles, Spanish-Language chains are in close proximity to major chains (in predominantly higher poverty tracts).

Figure 12: Central Los Angeles Major and Spanish-Language Supermarket Chains



In the second map (figure 13), red (dark grey in print) points indicate specialty stores and white points major chains. Specialty chains are geographically clustered in west Los Angeles and Santa Monica. Major chains in lower poverty tracts are in close proximity to specialty stores. Major chains in higher poverty tracts are geographically segregated from specialty stores.

Figure 13: Central Los Angeles Major and Specialty Supermarket Chains



Sensitivity Analyses

A series of sensitivity analyses compare the main analysis results to alternative store environment measures and stratification by racial/ethnic subgroups.

In the literature to date, most studies test the association of individual health behaviors or health outcomes with aggregated or ecologic measures of food store environments. To replicate these approaches similar ecologic measures are created, 1) a measure defined by whether the major chain existed in the respondents' home tracts[58], and 2) a measure defined by whether any store with annual sales greater than $2 million existed in the home tract.

As indicated previously about 15% of respondents live in tracts with a major chain store. In fully adjusted models of body mass index with the variable for major chain store included, it is a not significantly associated with BMI. For the variable constructed from large stores, about one-third of respondents live in tracts with large stores. In fully adjusted models of body mass index with the variable for large stores included, it is a not significantly associated with BMI.

Unlike earlier studies, the association with reported store types and BMI can be examined, as has been demonstrated in the store type analysis. To simplify the approach, the store type categories created can be collapsed into a single category of high frequency major chains. The stores included in this category based on a cut off frequency count of 20 stores in the InfoUSA 2009

[58] Identical to the measure created in part 1

database for Los Angeles County are: Ralphs, Vons, Albertsons, Food 4 Less, Trader Joe's, Whole Foods, Stater Brothers, and Vallarta. This frequency approach groups stores present in parts of four separate store type categories in the main analysis (major, discount, specialty and Spanish-language). In fully adjusted models of body mass index with this reported high-frequency major chain store included, shopping in these stores is associated with *higher* BMI (p = 0.051).

Because of geographic segregation among store types and among respondents of different racial/ethnic groups, stratified analysis tests whether the main analysis results are the result of similar associations in subgroup-store type combinations or averaging of different effects in observed combinations. This reflects the reality that for some subgroup-store type combinations there are few or no observations (see table 13 of store type frequency stratified by race/ethnicity).

Because of sample size limitations stratification analyses by Latino (n = 523) and white (n = 227) respondents are presented. Among Latino respondents the direction and magnitude of association between Spanish-Language chains and major chains (not stratified by poverty) is not statistically significant. Among white respondents the direction and magnitude of association between specialty stores and major chains (stratified by poverty) was similar to the main analysis sample and the p-value remained statistically significant (p = 0.001).

In addition to the question of fast food consumption, L.A. FANS wave 2 asked respondents about total daily servings of fruit and vegetables. Total servings for fruits, vegetables and a measure combining both is included as a sensitivity analysis to see if these measures of diet might mediate the relationship between store types and BMI. None of the variable constructs resulted in changes in the relationship between store types and BMI. One construct, dividing the sample by the median number of fruits and vegetables consumed, four servings per day, was statistically significantly associated with lower BMI.

The main analysis uses measured BMI from L.A. FANS wave 2 respondents. A similar sample was constructed from L.A. FANS wave 1 respondents using self-reported BMI as the dependent variable. The sample size for that group is 2297. In that sample the global F-test of store type is statistically significant.[59] The full model results of all sensitivity analyses are presented in Appendix G.

Conclusions

This is the second study (Drewnowski et al. 2012) in the United States to test the association between body mass index and reported food shopping behavior. It also tests the association for other food store types beyond the usual supermarket or chain supermarket dichotomy, for example discount, specialty or

[59] This is likely due to differences in sample size between the two analyses, 915 versus 2219.

Spanish-language chain stores. The results from the part 2 analyses suggest that shopping in some store types is independently associated with body mass index. The store type category created tests novel store categories as well as other store categories previously tested in the literature.

The major chain store category is different and more specific than other conceptions of this category in the prior literature, excluding specialty stores, Spanish-language stores and discount stores, which would have been grouped together in other studies. This store type category is divided by levels of poverty of the Census tract in which it is located, either by dividing the reported sample in half, or into additional categories, which test differences in the extremes reported in this sample. The findings suggest that there are no differences in these stores across levels of tract poverty in average body mass index after controlling for individual and household characteristics. These results differ from studies that suggest there are differences in health outcomes by when comparing supermarket store types to other stores types. The contents of stores are not measured in this study, but prior studies have suggested that store contents in these stores may differ in the types of food sold (Sloane et al. 2003). If these differences do exist then they may not translate into differences in body mass index.

The main differences observed between stores types in this analysis are between specialty food stores, Spanish-language chains and major chains in higher poverty tracts. In the cases of specialty food stores and Spanish-language chains, the direction of association with BMI is similar; both are associated with statistically significant lower BMI compared to major chains in higher poverty tracts. Specialty food stores are exclusively located in lower poverty tracts, similar to major chains in lower poverty tracts. They represent 5% of responses in the survey and the majority of respondents in this group were white. In stratified sensitivity analyses of white respondents the association between specialty food stores and BMI remains significant in the same direction of association. Since these results are cross-sectional the potential direction of association is unknown. It is possible that individuals shopping in these stores have unmeasured characteristics associated with BMI not accounted for by the individual covariates in the model, including the control for fast food consumption and exercise. Conversely, the stores could have characteristics that are associated with lower BMI assuming that the categories created are associated with differences in stores that are associated with lower BMI.

For those shopping in Spanish-language chain stores, BMI was lower on average than shoppers major chains stores in higher poverty tracts. The median poverty of the Census tracts in which these stores are located is 28% compared to 14.6% for major chain stores in higher poverty tracts. As the maps for each store chain show, these Spanish-language chains are located in a distinct geography from both specialty store and most major chain stores in higher poverty tracts. Latino shoppers make up the majority of shoppers in Spanish-language chain stores. In the stratified analysis by Latino ethnicity no store types are associated with BMI at a statistically significant level. The other major group shopping in Spanish-language stores is black respondents. The sample is too small for a stratified sensitivity analysis. Therefore it appears that it is not solely the relationship between Latinos and Spanish-Language chains that results the overall significant difference in the main analysis. Like the previous results, since these results are cross-sectional the potential direction of association is unknown. It is possible that individuals shopping in Spanish-language stores have unmeasured characteristics associated with BMI not accounted for by the individual covariates in the model, including the control for fast food consumption and exercise. Conversely, these stores could have characteristics that are associated with lower BMI assuming that the categories created are associated with differences in stores that are associated with lower BMI.

In prior work as discussed, some studies have observed differences in shopper BMI by discount store types, either measured by store name or directly through prices in stores (Drewnowski et al. 2012; Chaix et al. 2012). This study found no difference in discount stores in average shopper BMI compared to other stores types. In addition, there were no differences in average BMI between independent, bulk and small market shoppers compared to other store types.

In addition to the focus on major chains in the literature, the effect of small markets as a substitute for major chain market shopping where these chains may be deemed inadequate providers of healthy food or relatively inaccessible has been a focus of policy interventions. In this study, when asked the question as framed, requiring a single answer, very few respondents report shopping in a small market. There are several limitations to the survey question because it explicitly uses the term groceries (versus a more general term like food) and was unable to capture more than one response. These limitations aside, small markets appear to play a small role is grocery purchases as reported in this sample.

These results are helpful in understanding the relative nature of food store type shopping in Los Angeles County. Just over a third of respondents in this sample reported shopping in major chains. Thus, for this population, while being the largest share of the sample, it represents far less than half of the overall group. Spanish-language chain shoppers are close behind representing just over a quarter of all responses, followed by discount store shoppers. These results highlight the need to look beyond just a focus on major chains, or as suggested above, small markets as substitutes. In this sample, the substitutes for major chains are Spanish-language and discount food store types.

The proportion of shoppers in each store type varies by race/ethnicity. While the overall rate of shopping in major chains is 37%, the proportion among Latino respondents is 20% and the proportion among other respondents is similar, 55% for white, 50% for black, and 55% for other respondents. However in each group, the substitutes for major chains vary. For black respondents, a high proportion shop in either discount (21%) or Spanish-language (18%) store types. For white respondents a high proportion shops in specialty chain stores

(17%). For other respondents, bulk (13%) and small markets (14%) make up a large share of the remaining store types. Thus for each racial/ethnic group the substitutes for major chains differ, and for Latino respondents major chain shoppers are in the minority.

With known shopping location verified by external databases it was possible to calculate distances to the shopped store along the street network. The median distance traveled to store is 1.4 miles. This is considered longer than a typically walked distance. By racial/ethnic groups, the distance traveled by black and white respondents is the same, a median distance of 1.6 miles, compared to 1.3 miles for Latino respondents and 1.4 miles for other respondents. Thus the distances traveled to shop for groceries are similar for all groups and in multivariate models the distance traveled is not significantly associated with BMI.

Walking at 3 MPH it would take approximately 30 minutes to cover the median distance reported by respondents. About 7 in 10 households reported owning at least a single car, however it was unknown whether this car was used for food shopping. Among Latino respondents 60% report owning a car, and 75% of white, 70% of black and 72% of other respondents report owning a car. Car ownership is not significantly associated with BMI.

While distance and car ownership are not significantly associated with BMI, distance and car ownership could interact with store types or be a precursor for store type choice. Households without cars might still travel to stores in cars by sharing rides with other auto owners.

Smoking status, family size, and neighborhood poverty are the remaining parameters significantly associated with body mass index. Smoking is known to be associated with lower body weight thus the result is expected. Increased family size was associated with increased weight. It is not clear whether that might be due to associations with pregnancy, or other unobserved characteristics associated with increased family size relevant to eating or physical activity. Likewise the strong associations with neighborhood poverty suggest there may be numerous other processes not captured by the existing covariates, likely at multiple scales, associated with differences in BMI.

Chapter 5 – Research Part 3 – Store Change and Store Type Change

Research Question

This section of the analysis answers the research questions: How common is store type change over a 6-year period? What individual and neighborhood characteristics are associated with reporting store type change over a 6-year period? These questions are relevant for two reasons: 1) They directly address the policy question of whether opening or closing stores results in shopping behavior change, and 2) sets the framework for assessing the direction of causal association between body mass index and chain supermarket shopping, whether shopping in a chain supermarket results in lower BMI, or if lower BMI results in selection of chain supermarkets.

Summary of Findings

Most non-movers in L.A. FANS shopped in stores that are present at both wave 1 and wave 2 survey periods, and for those individuals rate of store type change is relatively low. For respondents that experience a store closure or opening, the rate of store type change increases substantially. In multivariate models this association remains.

Literature Review Summary

The goal of the literature review is to: 1) Describe how the literature has approached assessment of causal association between food environments and BMI, 2) Describe the findings from these studies. The attached table in Appendix

C (Part 3) summarizes the studies (Sturm and Datar 2005; Powell 2009; Gibson 2011; Leung et al. 2011; Block et al. 2011; Auchincloss et al. 2012). Store type change may be more important for BMI change than current measures of food environment change if, as suggested from the results of part 2, shopping in specific store types is associated with BMI difference.

One method used to assess the direction of causal association is to observe change in the dependent variable of interest while observing change in the independent variable of interest over a period of time. In this case, we might observe whether BMI increases when the number of chain supermarkets decreases over a period of several years while accounting for any other changes in individual characteristics, which may change at the same time and be related to BMI.

Six studies examined the longitudinal association between BMI and some measure of the food environment within the body of literature discussed previously. Three studies examined children/adolescents and three examined adults. Follow-up length ranged from 3 to 6 years in all but one study in which follow-up averaged 30 years. Of the three studies in children/adolescents none found an association between change in BMI and changes in food place measures, such as number of supermarkets or fast food restaurants.[60] For example one study, of girls only, found that convenience stores within a 0.25 mile network buffer of home were positively associated with BMI in cross sectional

[60] Two studies did find that food prices (for example for fruits and vegetables or fast food) were associated with BMI.

analysis but when examining change over 3 years the results were not statistically significant.

In adults, of the 3 studies, two found no association between change in BMI and change in measures of the food environment. One study found an association between BMI and food environment for specific populations. For example, for exclusively urban dwelling participants, the density of small markets was positively associated with BMI. For individuals that moved from rural to urban settings, the change in density of supermarkets, full-service restaurants, and small grocery stores was associated with changes in BMI, varying in direction depending on the food venue type.

Hypothesis

The main hypotheses for this section are: Opening and closing of stores stimulates a change in the store type of the respondent's store. A decrease in distance to the store shopped in at wave 2 compared to wave 1 results in switching of store types.[61]

Data

In part 3 I use data from L.A. FANS wave 1 and wave 2. L.A. FANS wave 2 re-contacted approximately 1,200 households from the original sample of 2,600

---

[61] The additional hypothesis, which is not testable here given data limitations, is whether switching to chain supermarket shopping from an alternative results in lower BMI.

households, roughly 6 years after their original survey date. The respondents completed identical survey questions in addition to several questions on diet and physical activity not asked in the first wave. The InfoUSA database from 2003 and 2009 is used to compare chain supermarket characteristics across years and assess store change.

Methods

The aim of the methodological approach is to assess the association between change in food store type shopping and fixed and changing individual and neighborhood characteristics that may be associated with this store type change.

Among the respondents to L.A. FANS participating in the wave 2 survey, one of 4 possible outcomes is possible (summarized in table 16):

Table 16: Matrix of Possible Outcomes from Moving and Food Shopping Place Type Change

	Did not Move	Moved
Shop in same store type	May have experienced change in local food environment, in situ	May have experienced change in local food environment by moving
Shop in different store type	May have experienced change in local food environment, in situ	May have experienced change in local food environment by moving

The analysis is limited to those that do not move over the 6-year period to make assessment of change in food environment over that period both tractable and considered exogenous.[62] In this case the primary influence on change in store type will be due to a change in the food environment[63] and change in household or neighborhood characteristics.

The primary outcome is change in store type of the respondent's store from wave 1 (t_0) to wave 2 (t_1). Comparing the store type between wave 1 and wave 2 and noting the change creates a dichotomous variable with no change as one category and any change as the other category. Thus the any change category contains many different store type changes as described later in the results.

The primary independent variable of interest is the opening or closing of stores, considered to be an exogenous influence on store type choice in non-movers. To define these stores, InfoUSA data from 2003 and 2009 are compared based on a unique identification number for each location. If the location was not present in the 2009 data it is classified as closed. If the location was not present in 2003 it was classified as a new store. This classification is assigned to each response based on the store matched to each response as described in earlier methods. For the small number of responses unmatched to the InfoUSA databases, change in store opening or closure was determined in a similar way. If any respondent reported a location at both waves it was considered to be open at both waves. If not, it was classified as a new opening if

---

[62] People who move over the 6-year period could select food environments with different characteristics with the intent to change store types.

[63] The difference between moving and in situ change is not trivial. Indeed, this one of the primary debates surrounding interpretation of results from the Moving To Opportunity (MTO) study.

present in wave 2 but not in wave 1, or a closure if present at wave 1 but not present at wave 2.

As discussed previously there are several methods for constructing environmental measures of food stores. The primary measure of environmental change is whether the number of stores in the home Census tract as measured by any store in the NAICS supermarket and grocery store category increased or decreased (i.e. any change). Other measures of environmental change considered are the presence of a chain supermarket, or distance to the nearest chain store. In addition to imputed environmental measures, differences in distance in the observed store selection (independent of store type change) can be assessed. Other covariates such as change in age, income, car ownership, household size, and neighborhood characteristics are included in models of store type change (summarized in table 17 below).

Table 17: Theoretical and Operationalized Measures Used in the Modeling Approach to Change in Store Type

Measure Class	Conceptual Measure	Operationalized Measure	Comment
Dependent variable	Change in typology of stores, or store attributes, that are either health promoting or detrimental to health	Change in store typed by industrial classification, size, and other characteristics hypothesized to be associated by health	Some evidence to support correlation between store type and health. Store type is based on a single report of store name and location as the most frequent location of grocery shopping.
Independent variable of interest	Exogenous influence on store type choice	Opening or closing of store in 6-year time	

Measure Class	Conceptual Measure	Operationalized Measure	Comment
Control variables – individuals	Mutable and immutable characteristics that may be independently associated with the dependent variable and if unaccounted for mask associations with the independent variable of interest	Age, Sex, Educational attainment Race/Ethnicity, Employment, Household Income, Nativity	These are derived from the prior literature, which has examined the associations between body mass index and individual characteristics.
Control variables – neighborhood	Mutable and immutable area level characteristics that may be independently associated with the dependent variable and if unaccounted for mask associations with the independent variable of interest	An index of educational attainment, family income, housing value, occupational status, and employment status.	These are derived from the prior literature, which has examined the associations between body mass index and area level characteristics.
Survey Structure	The survey oversampled specific populations which must be accounting for in the statistical analysis	Poverty Strata (Very Poor, Poor, Non-Poor); Households with children (Y/N); Survey weight	
Other variables – individual	Change in characteristics which may be associated with the independent variable of interest	Household automobile availability, Household size, Distance to store	

Measure Class	Conceptual Measure	Operationalized Measure	Comment
Other variables - neighborhood	Change in characteristics which may be associated with the independent variable of interest	Population density, local competition (distance between two closet large stores), proportion going to a specific type of store in the sample, change in distance to nearest chain supermarket	

Given the limitations of estimating survey weighted multilevel (MLM) multivariate logistic regression models, multivariate survey weighted logistic models with standard errors adjusted for clustering by Census tract are estimated. The dependent variable is change in store type and time varying individual and neighborhood characteristics are included in the model as differences between wave 1 and wave 2.

Model Example:

Model 2 – Change in food place type (no change, any change) = change in distance to store + change in time varying individual covariates + change in time varying neighborhood covariates[64] + survey structure covariates, given by

$$y_{ijt_0-t_1} = \beta_0 + \beta_n x_{nijt_0-t_1} + \beta_m x_{mjt_0-t_1} + (u_{0j} + e_{ij})$$

[64] Including change in ecologically imputed food environment measures.

where x_{nij} is a vector of n individual level predictors and x_{mj} is a vector of m tract level predictors.

Results

Store Type Change

The results of change in store type over 6-year follow up are presented. This analysis identifies the factors associated with change in store types such as the opening and closing of stores. The analysis is limited to 621 non-movers with reported stores in wave 1 and wave 2 of L.A. FANS and accounts for change in individual, household and local neighborhood characteristics.

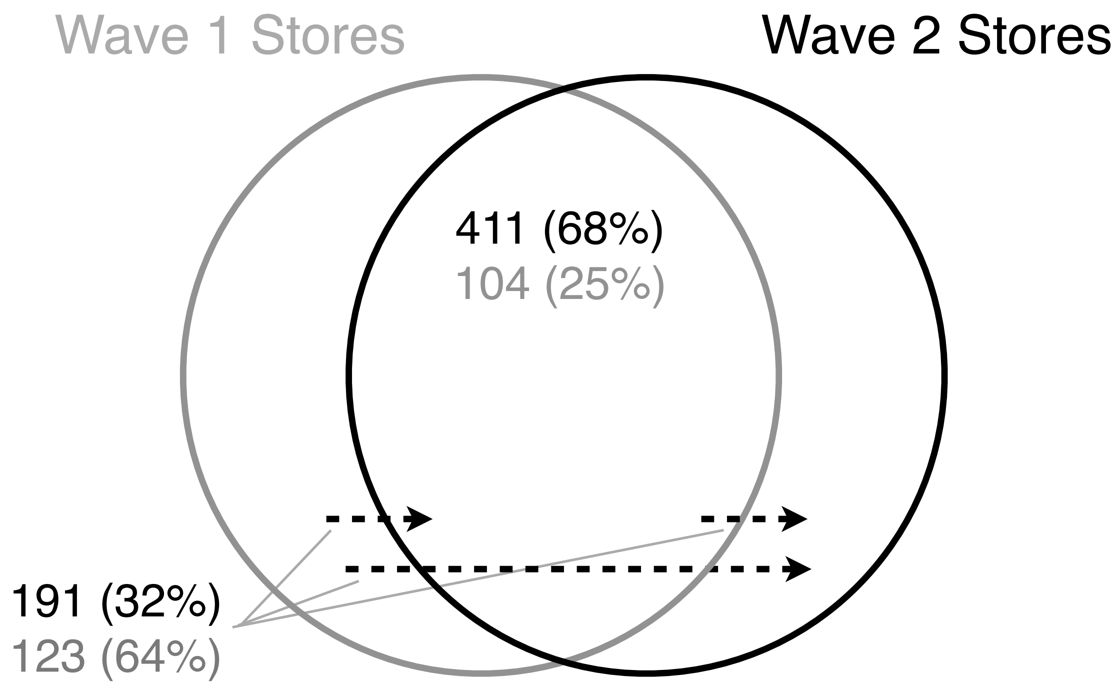
Table 18 below indicates the frequency of store type responses among non-movers between wave 1 and wave 2. The diagonal (shaded) indicates no change in store type between the waves. Overall, approximately 62% of non-moving respondents did not change store type over the 6-year follow-up. No change in store type is the dependent variable in the multivariate models (compared with any change, all other cells in the table).

Table 18: Matrix of Wave 1 Store Type Responses (rows) and Wave 2 Store Type Responses (columns)

	Discount	Specialty	Spanish Lang.	Major	Independent	Bulk	Small
Discount	10%	0%	4%	2%	0%	1%	1%
Specialty	0%	2%	0%	<1%	0%	<1%	0%
Spanish Lang.	2%	0%	12%	1%	<1%	<1%	<1%
Major	2%	4%	3%	34%	1%	3%	1%
Independent	1%	0%	3%	1%	2%	<1%	1%
Bulk	<1%	<1%	0%	1%	0%	1%	<1%
Small	1%	<1%	<1%	1%	1%	<1%	0.5%

Figure 14 below depicts the change in stores over a 6-year period and the frequency of change in store types in respondents that shopped in stores present at both waves compared to other respondents. Each circle in the figure represents the set of stores present in the InfoUSA database in 2003 (wave 1) and 2009 (wave 2). The majority of non-moving respondents, 411 (68%) shopped in stores present at both waves. Among that group, 104 (25%) changed the type of store shopped in over a 6-year period. In the remaining group, 191 (32%) respondents, the rate of store type change was higher 64%. This group was composed of respondents 1) where the wave 1 shopped store closed and the wave 2 store was open at both waves, 2) where the wave 1 shopped store closed and the wave 2 store opened in the 6-year period, or 3) where the wave 1 shopped store was open at both time periods and the wave 2 shopped store opened in the 6-year period.

Figure 14: Store Change and Store Type Change Over 6-year Follow-up



The frequencies in each category indicate that the rate of store type change is higher in non-movers who experience a change in the store in which they shop, with either that store closing or opening.

Multivariate models estimate the odds of store type change accounting for changes in individual, household and neighborhood characteristics. When accounting for these covariates, store opening and closings are significantly associated with store type change (OR 5.6, $p < 0.001$). In addition to store opening/closures, younger age, foreign born and change in income are associated with store type change. Also, most wave 1 store type categories are associated with store type change when major chain store type in higher poverty tracts is the reference category.[65]

Conclusions

This section assesses the association of store closures and openings on store type decisions, a relevant policy question since there are several existing programs, which aim to promote new store development.[66] This is also important because of the implied benefit this change may have on health outcomes or health behaviors.

Overall the store environment of most respondents is static as reported by shopping over a 6-year period. Most respondents shop in stores present at both time periods 6-years apart. Within that group the rate of store type change is low, about 1 in 4 change store types. This indicates that overall, store environments are relatively stable and that changes in store types are relatively rare. When stores open or close the rate of store type change increases. Among respondents who reported shopping in stores that closed or opened during the 6-year period over half changed store types indicating that store change may induce store type change.

In multivariate models the association between store type change and store openings and closing remains. In addition age, foreign born and change in

---

[65] Full model results are available in the Appendix H.

[66] For example in California the FreshWorks partnership (http://www.cafreshworks.com) and nationally the Healthy Food Financing Initiative, a partnership between the United States Treasury, Health and Human Services, and Department of Agriculture (http://www.acf.hhs.gov/programs/ocs/resource/healthy-food-financing-initiative-0)

income are associated with store change. These results indicate that store opening and closing may be helpful in stimulating store type change. However, whether specific store types changes are associated with changes in BMI is unknown. Also whether any specific direction of store type change is associated with BMI change is unknown. If specific store types are associated with higher or lower average BMI among shoppers in those types, then the direction of store type change is important for promoting BMI change in a positive direction. As indicated in the table 18 above many of the cells have very few or no responses. Thus using this study design it would be difficult to measure BMI change when specific store type changes occur.

One limitation of this analysis is that it does not divide store type change separately between openings and closings. This can be performed in future analysis, but as described earlier, there are several possible store opening and closing phenomena for a single respondent with two store measures separated over time. Another limitation to this analysis is the known rate of underlying store type change associated with store opening and closings. Store openings and closings could be associated with higher store type change because of competition or store type evolution which would promote different store types compared to adjacent or replacement stores.

Chapter 6 – Discussion

As stated in the introductory chapter, the goal of this research project is to provide new insights into the intersection of health and planning through the lens of a case study of shopping behavior and its relation to body mass index. This section discusses the major empirical findings, the limitations of the research, the implications of the research and recommendations for future research.

This research makes three contributions to the literature. In chapter three, a systematic test of the association between ecologically imputed measures and reported behavior is performed. This has not been systematically conducted in the literature to date. This systematic approach also allows for assessment of ecologically imputed measures and conclusions about whether their use is appropriate when considered against observed behavior.

In chapter four, the association between novel food store types, beyond the usual supermarket or chain supermarket dichotomy, and body mass index is tested. This is only the second study (Drewnowski et al. 2012) in the United States to test the association between body mass index and observed food shopping behavior.

In chapter five, a hypothetical policy question is tested; the effect of store closures or openings on store type decisions. This is important because of the implied benefit store type change may have on health outcomes or health behaviors.

Results from the third chapter show that in the 65 sampled 1990 Census tracts, only 7 contained a major chain supermarket as commonly defined based on data provided by InfoUSA. There are clear differences between poverty strata in the prevalence of chain supermarkets – there are none in very poor tracts. Conversely the rate of reported supermarket shopping in non-poor respondents is 84% despite only 17% of this group residing in a Census tract with a chain supermarket. These extreme examples demonstrate the limitation of this approach and the broad misclassification in both directions of assignment that may take place with the use of imputation.

Conceived as a medical test or screening tool, the knowledge that an individual lives in a tract with a chain supermarket may be associated with shopping in a chain supermarkets. However, the majority of the sample lives in tracts without chain supermarkets, and in this group, the test is of limited utility.

The measurement error modeling approach also contributes to understanding the limitation of imputation. This modeling approach, with reported chain shopping as the dependent variable and imputed chain shopping as the main independent variable of interest helps identify estimators, which may be biased. Because there are several predictors that are statistically significant, these results indicate that caution should be taken when models with dependent variables such as body mass index include both imputed chain shopping (presence of a chain store in the Census tract) along with these common individual and neighborhood characteristics.

Results from chapter four suggest that some store types are independently associated with body mass index. The formulated store type categories test novel store categories as well as other store categories previously tested in the literature.

The major chain store category is different and more specific than other conceptions of this category in the prior literature, excluding specialty stores, Spanish-language stores and discount stores, which would have been grouped together in other studies. Based on store location Census tract poverty, the major store type category is divided in half. The findings suggest that there are no differences among shoppers in these stores across levels of tract poverty in average body mass index after controlling for individual and household characteristics.

The main differences observed between store types in this analysis are between specialty food stores, Spanish-language chains and major chains in higher poverty tracts. In the cases of specialty food stores and Spanish-language chains, the direction of associated with BMI is similar; both are associated in statistically significant lower BMI compared to major chains in higher poverty tracts. Specialty food stores are exclusively located in lower poverty tracts, similar to major chains in lower poverty tracts. They represent 5% of responses in the survey and the majority of respondents in this group were white. In stratified sensitivity analyses of white respondents the association between specialty food stores and BMI remains significant in the same direction of association. Since these results are cross-sectional the potential direction of association is unknown. It is possible that individuals shopping in these stores have unmeasured characteristics associated with BMI not accounted for by the individual covariates in the model, including the control for fast food consumption and exercise. Conversely, the stores could have characteristics that are associated with lower BMI assuming that the categories created are associated with differences in stores that are associated with lower BMI.

For those shopping in Spanish-language chain stores, BMI is lower on average than major chains stores in higher poverty tracts. The median poverty of the Census tracts in which these stores are located is 28% compared to 14.6% for major chain stores in higher poverty tracts. As the maps for each store chain show, these Spanish-language chains are located in a distinct geography from both specialty store and most major chain stores in higher poverty tracts. Latino shoppers make up the majority of shoppers in Spanish-language chain stores. A stratified analysis by Latino ethnicity finds no association between any store types and BMI at a statistically significant level. The other major group shopping in Spanish-language stores is black respondents. The sample is too small for a stratified sensitivity analysis. Like the previous results, since these results are cross-sectional the potential direction of association is unknown. It is possible that individuals shopping in Spanish-language stores have unmeasured characteristics associated with BMI not accounted for by the individual covariates in the model, including the control for fast food consumption and exercise. For example, while the study does control for nativity and language of interview, there may be other unobserved factors, for example a broad concept such as acculturation, which may contribute to these findings. Conversely, these stores could have characteristics that are associated with lower BMI assuming that the categories created are associated with differences in stores that are associated with lower BMI. Again, for Spanish-Language chains there could be an element of acculturation by firms, in which they retain alternative selling practices, compared to other firms, which have positive associations with health.

In prior work as discussed, some studies have observed differences in BMI by discount store types, either measured by store name or directly through prices in stores. This study found no difference in discount stores in average BMI compared to other stores types. In addition, there were no differences in average BMI between independent, bulk and small markets compared to other store types.

In addition to the focus on major chains in the literature, the effect of small markets as a substitute for major chain market shopping where these chains may be deemed inadequate providers of healthy food or relatively inaccessible, has been a focus of policy interventions. In this study, when asked the question as framed, requiring a single answer, very few respondents report shopping in a small market.

These results are helpful in understanding the relative nature of food store type shopping in Los Angeles County. Just over a third of respondents in this sample reported shopping in major chains. Thus, for this population, while being the largest share of the sample, it represents far less than half of the overall group. Spanish-language chain shoppers are close behind representing just over a quarter of all responses, followed by discount store shoppers. These results highlight the need to look beyond just a focus on major chains, or as suggested above, small markets as substitutes. In this sample, the substitutes for major chains are Spanish-language and discount food store types.

The proportion of shoppers in each store type varies by race/ethnicity. While the overall rate of shopping in major chains is 37%, the proportion among Latino respondents is 20% and the proportion among other respondents is similar, 55% for white, 50% for black, and 55% for other respondents. However in each group, the substitutes for major chains vary. For black respondents, a high proportion shop in either discount (21%) or Spanish-language (18%) store types. For white respondents a high proportion shop in specialty chain stores (17%). For other respondents, bulk (13%) and small markets (14%) make up a large share of the remaining store types. Thus for each racial/ethnic group the substitutes for major chains differ, and for Latino respondents major chain shoppers are in the minority.

With known shopping location verified by external databases it is possible to calculate distances to the shopped store along the street network. The median distance traveled to store is 1.4 miles. This is considered longer than a typically walked distance. By racial/ethnic groups, the distance traveled by black and white respondents is the same, a median distance of 1.6 miles, compared to 1.3 miles for Latino respondents and 1.4 miles for other respondents. Thus the distances traveled to shop for groceries are similar for all groups and in multivariate models the distance was not significantly associated with BMI.

About 7 in 10 households report owning at least a single car, however it is unknown whether this car is used for food shopping. Among Latino respondents 60% report owning a car, and 75% of white, 70% of black and 72% of other respondents report owning a car. Car ownership is not significantly associated with BMI. While distance and car ownership are not significantly associated with BMI, distance and car ownership could interact with store types or be a precursor for store type choice.

Overall the store environment of most respondents is static as reported by shopping over a 6-year period. Most respondents shop in stores present at both time periods 6-years apart. Within that group the rate of store type change is low, about 1 in 4 change store types. This indicates that overall, store environments are relatively stable and that changes in store types are relatively rare. When stores open or close the rate of store type change increases. Among respondents who reported shopping in stores that closed or opened during the 6-year period over half changed store types indicating that store change may induce store type change.

In multivariate models the association between store type change and store openings and closing remains. In addition age, foreign born and change in income is associated with store change as well. These results indicate that store opening and closing may be helpful in stimulating store type change. However, whether specific store types changes are associated with changes in BMI is unknown. It is unknown whether simply any store type change is associated with BMI changes and whether the direction of change is important. Given many store type changes have very few or no responses using this study design it would be difficult to measure BMI change when store type changes.

Limitations

There are three types of limitations contained in the research presented here: 1) data limitations, 2) analysis limitations, 3) and external validity limitations given the focus is Los Angeles. Regarding data limitations, given the primary research question, assessment of the association between store types and body mass index, there are limitations within the dataset to capture important connections between the two concepts. The analytical method of the primary research question is cross-sectional. The study is focused in Los Angeles, which limits the applicability to other regions.

The primary limitation of this study is the nature of the survey question used to identify store types. It explicitly uses the term groceries (versus a more general term like food) and was unable to capture more than one response. Recent data from the U.S. Department of Agriculture indicate that in 2011, 42%

of food expenditures are for food consumed away from home.[67] Even if this is the case, multiple stores could be the source of this purchased food, and not just a single store. Another assumption is that store types are correlated with health promoting food, but food store contents are not directly measured. Also, frequency of shopping is unknown, which could be associated with BMI in multiple pathways.

In addition to these limitations, the analysis of store types and body mass index is cross-sectional. It is impossible to infer the direction of association between store types and body mass index. It is possible that unobserved individual characteristics produce the observed association between store types and BMI. It may also be possible that there are characteristics of stores that contribute to lower BMI in individuals. Likewise, there could be mixing of effects in a single store, masking some effects in store types observed to have no effect, or masking larger effects, or null effects in some individuals shopping within store types with a positive association with BMI. Despite these limitations this work has addressed important limitations in the prior literature, filled gaps in the literature and made a contribution to advancing the understanding of how food stores may influence health. Table 19 below summarizes some of these contributions.

---

[67] Table 10 "Food away from home as a share of food expenditures" (dated 10/1/2012) USDA Economic Research Service Food Expenditure Series (available at: http://www.ers.usda.gov/data-products/food-expenditures.aspx - accessed: 5/28/2013)

Table 19: Summary of Prior Research, Gaps and Contribution to the Literature

Prior Research	Gap(s)	Contribution(s)
• Ecologic/Geographic Association	• Environmental determinism • Ecologic fallacy • Geographic scale and travel	• Comparison of ecological imputation to reported shopping • Distance to reported store and car availability
• Supermarket focus • Supermarket and SES or race interaction	• Heterogeneity in supermarkets beyond SES • Other racial/ethnic groups	• Store types • Large Latino sample in L.A. FANS
• Cross-sectional or Longitudinal with ecologic association	• Unknown how store change is related to outcomes	• 6-year longitudinal follow-up for store type outcome

Implications for Theory and Policy

Despite the above limitation, the empirical findings nonetheless have theoretical and policy implications. Environments exert powerful forces on objects within their scope of influence. Darwin described the powerful relationship between natural environments and organisms experiencing random genetic variation and reproductive pressure. But this powerful metaphor may not be the best approach to thinking about how individuals procure food and whether that spatial activity influences health. Rather, individuals navigate economic markets of food selling firms in which both have agency. Individuals are likely constrained by external forces that cluster groups in common locations, and because location is one form of market power available to food selling firms by nature of the monopolistic competition present in the market, individuals may be subject to this market power to the detriment of their health. It is this conceptual framework, which provides a path forward to understanding the connection between consumers, food sellers and health.

If the food industry is an example of monopolistic competition then this has implications for policy. The main results suggest that shopping in specialty and Spanish-language store types is associated with lower BMI, compared to the referent category major chain stores in higher poverty tracts. If these three types are considered differentiated by health as demonstrated empirically then the theory suggests that this could be because of three reasons: type, location, or quality. Looking at the map it can be seen that in general the store types are differentiated by location, i.e. they are in different places. They may be differentiated by quality, but that was not directly measured. However if we consider how stores could be differentiated by health then quality could be causally linked to differences in health. Finally, they could be differentiated by type, in the sense that different stores have different characteristics recognized at types (think brands or “organic”, versus Spanish, versus “the major chain”). These type identities could possibly be linked to health, but probably not in the intended direction of association if type examples include “healthy.” As just described, in the case of these highlighted store types, location may be the strongest differentiating factor yet the weakest link to health, at least causally. Thus we observe associations, which may in fact be correlations with health rather than causal associations. But again, because location may also be correlated with things like quality and type, it makes it *appear* as these might be the factors associated with health rather than just a spurious association.

This poses a challenge to policy makers because the desirable outcome of an association between shopping in a store and health is not what it appears to be. Rather it is the clustering of individuals by larger social forces and the responses of firms to this outcome, which results in the outcome. Thus the policy intervention becomes altering the level of location differentiation available to the market, or addressing the level of demand that exists within given locations, or directly addressing spatial interactions so as not to allow location to be such an important differentiator.

As discussed, the main results from this analysis show that there are differences in shopper BMI by specific store types. This is a cross-sectional analysis so the direction of causation is unknown. The association could be due to unobserved characteristics of individuals unaccounted for by the control variables. The association could also be due to characteristics of the store that are associated with differences in BMI. The outcome could be due to a combination of both, present within and across individuals shopping in a specific store.

If shopping in specific store types is associated with health in the direction of stores independently influencing heath, then the results suggest more attention should be paid to specific types and brands rather than larger categories based on industrial classification or sales volume alone. Therefore policies that encourage these specific stores in specific places could improve health outcomes.

If the results are due to unobserved characteristics that are simply captured by location differentiation then the approach of promoting specific types many not result in health benefit. Alternatives may be to limit location differentiation, however it is unclear how that might be associated with health. The implication of unobserved characteristics is that further research has to be done to identify these characteristics. Also, because larger structural measures like the level of tract poverty continued to be associated with BMI these factors also have to be measured and addressed. Changing the characteristics of individuals in their interaction with the firms that sell food may be another possible intervention.

Future Work

Future analysis grouping specialty stores and major chains would help to identify whether prior results attributed to major chains are in fact driven by the lack of segmenting the shopper BMI association into finer store types as done here. A sensitivity analysis grouping Latino and black respondents to test the association with Spanish-language chains would capture the majority of the sample that shops in Spanish-language chains.

Of major importance is accounting for the potential for reverse causation – the association of store type and shopper BMI based on unobserved individual characteristics. This will likely have to come from observation of natural experiments following change in stores over time in conjunction with change in health behaviors and health outcomes. Studies will have to account for location, type and quality as differentiators along with the characteristics and constraints of individuals. This can be undertaken because travel to food stores can be used as a marker of preference and at the same time, assess the role of location, mediators of location, and in many cases also can be used as a proxy for type and quality assuming that exact locations are known.

Because this work focuses on a spatial question the relevance of location measurement is readily apparent. This work also highlights the complex causal relationships that may be present and not readily addressable without rethinking how to design empirical studies to evaluate these relationships. Fortunately we live in a new era with the ability to measure location ubiquitously and at a large scale. For example in the United States, approximately 100 million adults[68] carry smartphones that permit highly accurate location measurement which can be captured continuously and remotely shared to a third party observer in real-time or intermittently.[69]

The tools of social science data collection have not changed markedly since inception of the science itself. They consist of direct observation, survey instruments, or use of administrative data. Methods for data analysis have

---

[68] Duggan, M. Cell Phone Activities 2012. Pew Internet & American Life Project, November 25, 2013, http://www.pewinternet.org/~/media//Files/Reports/2012/PIP_CellActivities_11.25.pdf accessed on April 9, 2013.

[69] http://www.fcc.gov/guides/wireless-911-services accessed on April 10, 2013 FCC - Federal Communication Commission

evolved in many ways and are the primary drivers of new inference, in addition to expanding the range of data collected using current methods. However, limitations of social science data collection do limit the nature of scientific inference. Collecting data over time, which assists causal inference by the nature of inter-temporal change, increases cost and difficulty. Other methods of causal experimentation are limited because of the difficulty in creating experimental conditions in the real world. Model systems and detailed measurement, the tools in which the natural and physical sciences most often employ, are simply more difficult for social scientists.

But with a new era in measurement possible, many of these limitations can be addressed. Passive location measurement has a very low burden for the user. In fact it could be argued that for 100 million adults today, the cost is zero. Combined with maps, and some additional individual data collected at a single point in time, the construction of measures that can be updated at very short temporal intervals and over long follow-up periods is now possible. Likewise, this level of longitudinal observation allows for the observation of natural experiments, the counterpart to model systems and randomized controlled trials present in the natural sciences. This detail of measurement or observation is also comparable to the instrumentation and observation common in natural and physical sciences.

Diet-related disease is highly prevalent in the United States. It is not evenly distributed in the population, diet-related outcomes like BMI differing by sex, race/ethnicity and other measures of social structure. In the face of these pressing health problems and a mixed picture as to the cause of these differences, interventions have to be developed both within and beyond typical policy frameworks. From this work, interventions that promote specific store types in specific places may be helpful. However, as also suggested by this work, in the face of unchanging social structure, the risk of this social structure must be reduced. For diet, this will likely come from detailed understanding of how social relationships influence food consumption and then explicitly intervening in those relationships. One important relationship will likely be between the individual and the food selling firm. Intervening in that relationship, always in part a spatial relationship, as a way to promote health may be an important next step in addressing health inequality from a spatial perspective.

Appendix A – Los Angeles Family and Neighborhood Survey Sample and Data Limitations

As with any data analysis, the results are limited by the set of survey questions available and the validity of the final sample in relation to a representative sample of the population.

Table 20 below indicates the relevant measures available for analysis in L.A. FANS for wave 1 and wave 2. The decision to focus the cross-sectional analysis on body mass index in wave 2, despite the smaller sample size (discussed below), reflects the fact that it was directly measured, removing self-reporting bias,[70] and also allows for control by known factors which would be associated with BMI such as physical activity and fast food consumption. In addition it allows for testing of a potential moderator of the relationship between store types and BMI, fruit and vegetable consumption, or in future analysis, the testing of fruit and vegetable consumption as a dependent variable itself.

Table 20: Measures available for analysis in L.A. FANS wave 1 and wave 2

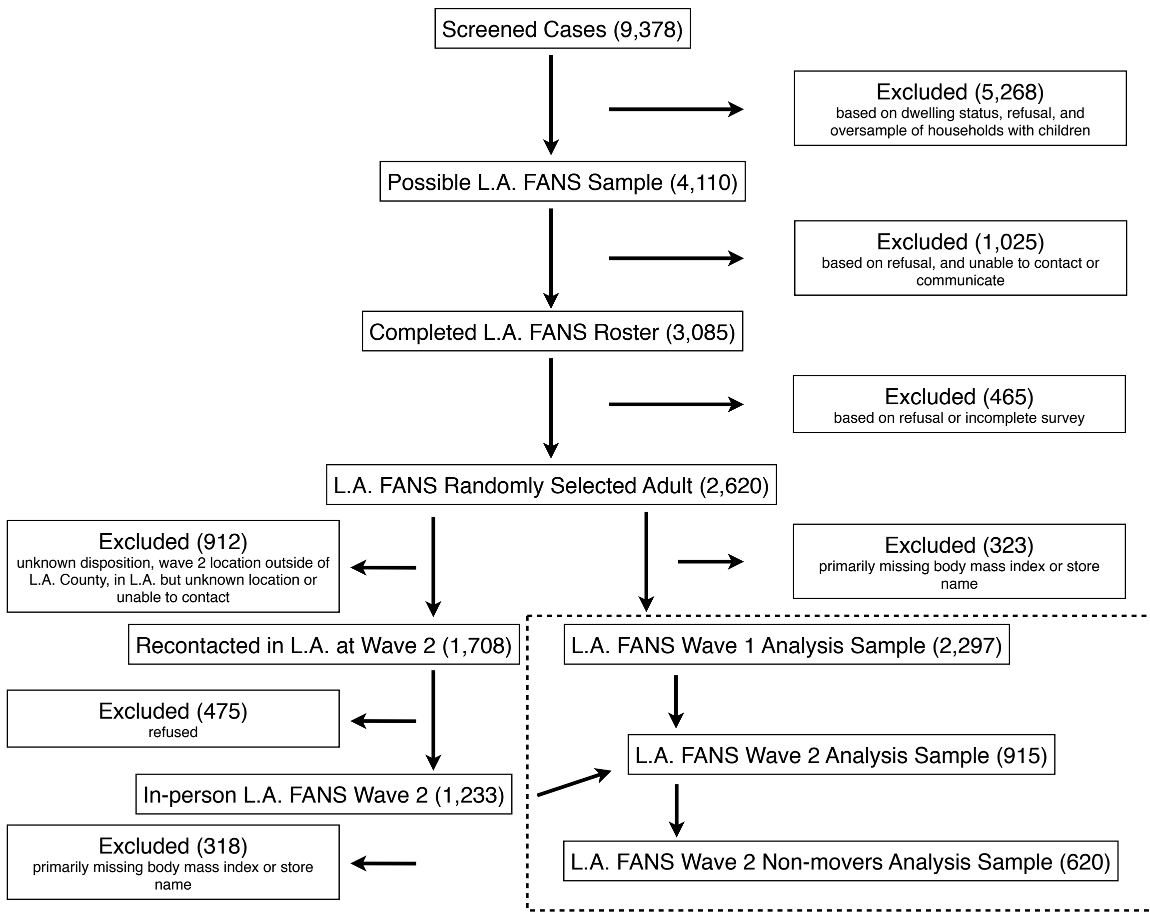
Measure	L.A. FANS Wave 1	L.A. FANS Wave 2
Body mass index (self-reported)	X	X
Body mass index (directly measured)		X
Store name and cross streets	X	X
Household car ownership	X	X
Fruit and vegetable consumption in the past 24 hours		X
Fast food consumption in the past 24 hours		X
Moderate or vigorous physical activity in the past week		X

[70] Whether this bias might be correlated with store name or type is unknown, but using self-reported BMI removes this possibility.

Figure 15 below outlines the selection of partipants from L.A. FANS during both wave 1 and wave 2 selection. The dotted-line box highlights the main analyses samples used in the three parts of this dissertation. The first, the wave 1 sample of 2,297 is derived from the main randomly selected adult sample who were asked the question related to grocery shopping. In most cases, excluded participants did not report body mass index or store name. Approximately six years later L.A. FANS followed up with wave 1 participants in Los Angeles County, repeating the survey in-person with 1,233 of the original wave 1 sampled adults. From this group, 915 are included in the main analysis of directly measured body mass index, having complete BMI and store name data.

Within the 915 wave 2 participants, 620 did not move during the six year period. This sample was used to assess the relationship between store openings and closing, and store type change over a six-year period.

Figure 15: L.A. FANS wave 1 and wave 2 sample selection



The tables below (table 21 and 22) compare the wave 1 characteristics of main wave 2 sample to the sample from wave 1 excluded from the sample for any reason, either unknown disposition, refusal to participate or exclusion due to missing data. Age, gender and racial/ethnic composition were similar between the two groups. Body mass index (based on self report in wave 1) and distance to store were similar between the two groups.

Table 21: L.A. FANS Wave 1 Characteristics Compared Between Analysis Sample and Excluded Participants (1)

	Wave 2	Wave 1
N (%)	915 (35%)	1704 (65%)
BMI* - kg/m2, mean (SD)	26.9 (5.2)	26.5 (5.2)
Distance to store - miles, MD (IQR)	1.14 (0.57-1.9)	1.11 (0.57-1.9)
Age - years, mean (SD)	39.7 (12.8)	39.7 (15.3)
Female	61%	58%
Race/Ethnicity		
Latino	58%	56%
White	25%	25%
Black	9%	10%
Other	8%	8%

In the wave 2 sample compared to the excluded wave 1 sample educational attainment was slightly higher, as was median household income and household car ownership.

Table 22: L.A. FANS Wave 1 Characteristics Compared Between Analysis Sample and Excluded Participants (2)

	Wave 2	Wave 1
Educational Attainment		
Less Than HS	35%	37%
High School	18%	21%
Some College	25%	24%
College or Higher	21%	17%
Household Income -$1,000 median (IQR)	30 (15-58)	25 (13-50)
Employed	68%	62%
Smoker	13%	16%
US Born	46%	45%
Married/Partner	65%	59%
Own Car	80%	72%

Comparing wave 1 store type frequency responses between the wave 2 sample and the excluded wave 1 sample (table 23 below), the frequency of store type responses is similar across types. The wave 2 responses from the wave 2 sample do show shifting in store type choices over the 6-year period. In that time the proportion shopping in specialty, Spanish-language, and bulk store types increases, and the proportion shopping in discount, independent and major chains decreases.

Table 23: L.A. FANS wave 1 (included and excluded) and wave 2 store type frequency

	Wave 1	Wave 2 (wave 1 resp.)	Wave 2 (wave 2 resp.)
N	1627	915	912
Discount	308 (19%)	186 (20%)	152 (17%)
Specialty	30 (2%)	22 (2%)	48 (5%)
Spanish L.	316 (19%)	180 (20%)	243 (27%)
Major	717 (44%)	398 (44%)	333 (37%)
Major (high SES)			160 (18%)
Major (low SES)			173 (19%)
Independent	139 (9%)	74 (8%)	40 (4%)
Bulk	46 (3%)	20 (2%)	63 (7%)
Small	71 (4%)	35 (4%)	33 (4%)

One aim of the L.A. FANS researchers was to oversample by poverty. This meant collecting data from tracts identified at “very poor,” “poor” and “non-poor” as defined by percentile Census tract poverty rank. Table 24 below outlines the estimated total population in each group of tracts for Los Angeles County in the year 1997 when the analysis was conducted. In each group the average percent in poverty was 47%, 30%, and 10% respectively.

Table 24: L.A. FANS original sampling plan and L.A. County population estimates

	Percentile pov. rank	Pct. pov.	Tracts	Pop	Pct. pop.
Very Poor	90-100	47%	161	881,956	9%
Poor	60-89	30%	490	3,302,831	34%
Non-Poor	1-59	10%	973	5,409,384	56%

* Adapted from table 2.1 L.A. FANS wave 1 codebook

From the very poor and poor strata of poverty an approximately equal proportion (30%) of the sample was to be obtained and the remainder (40%) was to come from the non-poor strata. In table 25 below, the first column indicates the estimated sample target, followed by the actual wave 1 sample obtained, stratified by poverty, and the proportion in the main wave 2 analysis sample and the excluded group. The wave 2 sample retains the stratification by poverty intended in the original design of the survey.

Table 25: L.A. FANS poverty strata in wave 1 and wave 2 samples

	Estimated sample population	Sampled	Wave 2	Wave 1
N (%)		2619 (100%)	915 (35%)	1704 (65%)
Very Poor	27%	30%	28%	31%
Poor	37%	31%	31%	31%
Non-Poor	36%	39%	41%	38%
HH with children		76%	80%	74%

Appendix B – Measurement Error Models

Epidemiologists generally assume that measurement error in an independent variable (x) is not correlated with the true value of the variable and hence the error biases coefficient estimates toward the null (Wacholder 1995).[71] However, if this is not the case, and the error systematically varies by the measured value,[72] then regression coefficients can be biased, resulting in spurious conclusions. To correct for measurement error, a validation study uses the "gold standard" exposure as the dependent variable in a model with the measured exposure and other covariates from the main model estimating the outcome of interest (Spiegelman 2010). When the gold standard is unavailable, results from the validation study are used to adjust regression coefficients in the main outcomes model.

Confounders may also have measurement error. While in general, even with measurements error, inclusion of a confounder in a regression model will result in less biased coefficient estimates, it is possible that with significant measurement error, especially in continuously measured confounders, and with correlation in error between exposure of interest and confounders, coefficient

---

[71] This is called the "classical error model."

[72] For example, if the error in estimating whether someone shops in a supermarket is larger for estimates of not shopping in a supermarket, compared to the error in estimating shopping in a supermarket.

estimates could be more biased with the confounder than without it (Wacholder 1995).[73]

[73] Wacholder concludes: "Indeed, errors that strongly correlate with the true value of the confounder or with the exposure can produce the apparent anomaly that adjustment for a poorly measured variable yields an estimate that is more biased than the crude." (Wacholder 1995) p. 160.

Appendix C – Literature Review Table

Table 26: Literature Review - Part 1 Studies - Citation and Measure Description

Table Id	Diss. Part	Pubmed ID	Author, Journal, Year	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
1	1	11777675	Morland K, Wing S, Diez Roux A, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. Am J Prev Med 2002 Jan;22(1):23-9.	511	population, segregation, race, income	count per tract (prevalence ratio)	tract	
2	1	15450633	Block JP, Scribner RA, DeSalvo KB. Fast food, race/ethnicity, and income: a geographic analysis. Am J Prev Med 2004 Oct;27(3):211-7.	247	population, race, income	density per buffer area	0.5, 1 mile buffers, around tract centroids	
3	1	15798127	Zenk SN, Schulz AJ, Israel BA, James SA, Bao S, Wilson ML. Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. Am J Public Health 2005 Apr;95(4):660-7.	240	population, race, income, poverty, segregation	distance from tract centroid to nearest chain supermarket	tract	

Table Id	Diss. Part	Pubmed ID	Author, Journal, Year	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
4	1	16530621	Morland K, Diez Roux AV, Wing S. Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. Am J Prev Med 2006 Apr;30(4):333-9.	212	individual, adult, cardiovascular risk factors, obesity, overweight, diabetes, cholesterol, hypertension, ARIC	count (presence) in tract	tract	
5	1	16380567	Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. Am J Public Health 2006 Feb;96(2):325-31.	210	population, race, income	counts in tract, population adjusted	tract	
6	1	16997358	Powell LM, Slater S, Mirtcheva D, Bao Y, Chaloupka FJ. Food store availability and neighborhood characteristics in the United States. Prev Med 2007 Mar;44(3):189-95.	181	population, race, income, SES	count per ZIP, population adjusted	ZIP	

Table Id	Diss. Part	Pubmed ID	Author, Journal, Year	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
7	1	15333313	Horowitz CR, Colson KA, Hebert PL, Lancaster K. Barriers to buying healthy foods for people with diabetes: evidence of environmental disparities. Am J Public Health 2004 Sep;94(9):1549-54.	176	store content, population, race, income, diabetes	food types in stores by neighborhood	neighborhood	

Table 27: Literature Review - Part 1 Studies - Summary of Results

Table Id	Food Place Types	Summary of Results	Assessment of Results	Geographic Location	Location Notes
1	supermarket, convenience, alcohol	Four times more supermarkets in white neighborhoods compared to black neighborhoods. More supermarkets and gas stations with convenience stores in wealthier neighborhoods compared with poorer neighborhoods. Three times more places for alcohol consumption in poorer neighborhoods compared to wealthier neighborhoods.	positive	Mississippi, North Carolina, Maryland, Minnesota	
2	fast_food	Density of fast food in 1 mile buffer was associated with median household income and percent black residents (accounting for commercial activity, presence of highways, median home values). Similar results in 0.5 mi buffer. In predominantly black neighborhoods there are 2.4 fast food restaurants per square mile compared to 1.5 in predominantly white neighborhoods.	positive	New Orleans, LA	156 tracts
3	supermarket_chain	Distance to the nearest supermarket was similar in least poor tracts independent of racial composition, in the most poor tracts predominantly African American tracts were 1.1 mile farther from chain supermarkets than white neighborhoods. Models were adjusted for spatial autocorrelation.	positive	Detroit, MI	869 tracts
4	supermarket, convenience	Presence of supermarkets in the tract was associated with lower prevalence of obesity and overweight, convenience stores were associated with a higher prevalence. Associations with diabetes, cholesterol, and hypertension were not observed.	mixed	Mississippi, North Carolina, Maryland, Minnesota	207 tracts
5	supermarket, grocery, alcohol, fruit_vegetable, bakery, specialty, natural_food	Minority and mixed neighborhoods had twice as many grocery stores than predominantly white, and half as many supermarkets. Poor and non-white neighborhoods had fewer alternative stores. Alcohol stores were more common in poor neighborhoods.	positive	New York, North Carolina, Maryland	

Table Id	Food Place Types	Summary of Results	Assessment of Results	Geographic Location	Location Notes
6	supermarket_chain, supermarket, grocery, convenience	Low-income ZIPs have 75% of chain supermarkets as middle-income ZIPs, African American ZIPs have 52% of white ZIPs, Hispanic ZIPs have 32% of non-Hispanic ZIPs. Models control for income and other covariates. Non-chain supermarkets and grocery stores are more prevalent in low income and minority ZIPs.	positive	United States	
7	grocery, bodegas	In East Harlem, 18% of grocery stores stocked recommended foods for diabetics, compared to 58% in the Upper East Side. For bodegas, 9% compared to 48%, stocked all items. There were more bodegas in East Harlem, and it was more likely to have stores on the block to no stock items and to have stores that did.	positive	New York, NY	East Harlem compared to Upper East Side

Table 28: Literature Review - Part 2 Studies - Citations and Measure Description

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
8	2	17884578	Powell LM, Auld MC, Chaloupka FJ, O'Malley PM, Johnston LD. Associations between access to food stores and adolescent body mass index. Am J Prev Med 2007 Oct;33(4 Suppl):S301-7.	116	adolescent, bmi, individual			
9	2	17496257	Wang MC, Kim S, Gonzalez AA, MacLeod KE, Winkleby MA. Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. J Epidemiol Community Health 2007 Jun;61(6):491-8.	100	individual, bmi, adult	area adjusted density by tract, distance	tract	
10	2	16777537	Inagami S, Cohen DA, Finch BK, Asch SM. You are where you shop: grocery store locations, weight, and neighborhoods. Am J Prev Med 2006 Jul;31(1):10-7.	92	individual, LAFANS, transportation, adult	distance, relative tract position from home	tract	

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
11	2	17712130	Lopez RP. Neighborhood risk factors for obesity. Obesity (Silver Spring) 2007 Aug;15(8):2111-9.	48	individual, BRFSS, bmi, obesity	(area?) adjusted density by ZIP, presence of supermarket in ZIP	ZIP	
12	2	19337520	Rundle A, Neckerman KM, Freeman L, Lovasi GS, Purciel M, Quinn J, Richards C, Sircar N, Weiss C. Neighborhood food environment and walkability predict obesity in New York City. Environ Health Perspect 2009 Mar;117(3):442-7.	42	individual, adult, land_use, bmi, walkability	area adjusted density in buffer	half mile buffer around exact home address	
13	2	19231301	Powell LM, Bao Y. Food prices, access to food outlets and child weight. Econ Hum Biol 2009 Mar;7(1):64-72.	36	children, bmi, individual, price, dunn_bradstreet	area adjusted, population adjusted	?	
14	2	19666158	Rose D, Hutchinson PL, Bodor JN, Swalm CM, Farley TA, Cohen DA, Rice JC. Neighborhood food environments and Body Mass Index: the importance of in-store contents. Am J Prev Med 2009 Sep;37(3):214-9.	28	bmi, adult, transportation	shelf space within 1km buffer of home	1km buffer from home	

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
15	2	19533365	Inagami S, Cohen DA, Brown AF, Asch SM. Body mass index, neighborhood fast food and restaurant concentration, and car ownership. J Urban Health 2009 Sep;86(5):683-95.	25	bmi, LAFANS, adult, transportation	density (type?) in tract?	tract	
16	2	19560992	Galvez MP, Hong L, Choi E, Liao L, Godbold J, Brenner B. Childhood obesity and neighborhood food-store availability in an inner-city community. Acad Pediatr 2009 Sep-Oct;9(5):339-43.	25	bmi, children	count (presence) in block	block (census)	
17	2	19766372	Zick CD, Smith KR, Fan JX, Brown BB, Yamada I, Kowaleski-Jones L. Running to the store? The relationship between neighborhood environments and the risk of obesity. Soc Sci Med 2009 Nov;69(10):1493-500.	16	bmi, obesity, dunn_bradstreet, adult	count (presence) in neighborhood	neighborhood	
18	2	20106710	Black JL, Macinko J, Dixon LB, Fryer GE Jr. Neighborhoods and obesity in New York City. Health Place 2010 May;16(3):489-99.	13	bmi, obesity, adult, population	availability		

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
19	2	18483833	Brown AF, Vargas RB, Ang A, Pebley AR. The neighborhood food resource environment and the health of residents with chronic conditions: the food resource environment and the health of residents. J Gen Intern Med 2008 Aug;23(8):1137-44.	8	bmi, adult, LAFANS, individual	roadway mile density in tract	tract	
20	2	21547411	Ford PB, Dzewaltowski DA. Neighborhood deprivation, supermarket availability, and BMI in low-income women: a multilevel analysis. J Community Health 2011 Oct;36(5):785-96.	1	adult, women, SNAP, bmi	count (presence) in neighborhood	neighborhood	
21	2	21563722	Gregson J. Poverty, sprawl, and restaurant types influence body mass index of residents in California counties. Public Health Rep 2011 May-Jun;126 Suppl 1:141-9.	0	bmi, population	ratio to total retail	county	

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
22	2	21660076	Dubowitz T, Ghosh-Dastidar M, Eibner C, Slaughter ME, Fernandes M, Whitsel EA, Bird CE, Jewell A, Margolis KL, Li W, Michael YL, Shih RA, Manson JE, Escarce JJ. The Women's Health Initiative: the food environment, neighborhood socioeconomic status, BMI, and blood pressure. Obesity (Silver Spring) 2012 Apr;20(4):862-71.	0	bmi, individual, women, adult, WHI	availability in buffer divided by percentile (10th compared to 90th)	1.5 mile buffer around home	
23	2	16436207	Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? Int J Behav Nutr Phys Act 2006;3:2.		bmi, adult, individual	proximity to home, work	home address	work address
24	2	19295936	Millstein RA, Yeh HC, Brancati FL, Batts-Turner M, Gary TL. Food availability, neighborhood socioeconomic status, and dietary patterns among blacks with type 2 diabetes mellitus. Medscape J Med 2009;11(1):15.		bmi, adult, individual	count (presence) in tract	tract	

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
25	2	20529405	Laska MN, Hearst MO, Forsyth A, Pasch KE, Lytle L. Neighbourhood food environments: are they associated with adolescent dietary intake, food purchases and weight status? Public Health Nutr 2010 Nov;13(11):1757-63.		adolescent, bmi, individual	availability within buffer, distance	800, 1600m, 3000m network buffer around home	school
26	2	20689393	Jilcott SB, McGuirt JT, Imai S, Evenson KR. Measuring the retail food environment in rural and urban North Carolina counties. J Public Health Manag Pract 2010 Sep-Oct;16(5):432-40.		population, bmi, measure_comparison	ratio, population density		
27	2	21272422	Casagrande SS, Franco M, Gittelsohn J, Zonderman AB, Evans MK, Fanelli Kuczmarski M, Gary-Webb TL. Healthy food availability and the association with BMI in Baltimore, Maryland. Public Health Nutr 2011 Jun;14(6):1001-7.		adult, individual, bmi	availability of healthy food measured by NEMS in tract	tract	

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
28	2	21486525	Jilcott SB, Wade S, McGuirt JT, Wu Q, Lazorick S, Moore JB. The association between the food environment and weight status among eastern North Carolina youth. Public Health Nutr 2011 Sep;14(9):1610-7.		bmi, individual, adolescent	count in buffer, distance to nearest	0.25, 0.5, 1, 5mi buffer (Euclidean and network) from home	
29	2	21551382	Hickson DA, Diez Roux AV, Smith AE, Tucker KL, Gore LD, Zhang L, Wyatt SB. Associations of fast food restaurant availability with dietary intake and weight among African Americans in the Jackson Heart Study, 2000-2004. Am J Public Health 2011 Dec;101 Suppl 1:S301-9.		bmi, adult, Jackson_Heart	availability in buffer around home	home address	
30	2	21852626	Keegan TH, Hurley S, Goldberg D, Nelson DO, Reynolds P, Bernstein L, Horn-Ross PL, Gomez SL. The association between neighborhood characteristics and body size and physical activity in the California teachers study cohort. Am J Public Health 2012 Apr;102(4):689-97.		adult, women, obesity			

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
31	2	22496738	Chaix B, Bean K, Daniel M, Zenk SN, Kestens Y, Charreire H, Leal C, Thomas F, Karusisi N, Weber C, Oppert JM, Simon C, Merlo J, Pannier B. Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). PLoS One 2012;7(4):e32908.		bmi, adult, Paris	observed shopping location	home	
32	2	22516505	Wall MM, Larson NI, Forsyth A, Van Riper DC, Graham DJ, Story MT, Neumark-Sztainer D. Patterns of obesogenic neighborhood features and adolescent weight: a comparison of statistical approaches. Am J Prev Med 2012 May;42(5):e65-75.		bmi, adolescent, individual			

Table 29: Literature Review - Part 3 Studies - Summary of Results

Table Id	Food Place Types	Summary of Results	Assessment of Results	Geographic Location	Location Notes
8	supermarket_chain, supermarket_nonchain, convenience, grocery	More chain supermarkets were associated with lower BMI	positive	United States	
9	grocery, grocery_small, supermarket_chain	Higher density of small grocery stores was associated with higher BMI among women, closer proximity to chain supermarkets was associated in higher BMI among women	mixed	California	Agricultural regions
10	grocery	[difficult to interpret]	mixed	Los Angeles CA	
11	retail, establishment, supermarket, fast_food	Population density, employment density, establishment density, presence of supermarket was associated with obesity risk (BMI > 30), fast food density was not associated with obesity risk	mixed	United States	
12	supermarket, fruit and vegetable markets, natural food stores	Higher density of supermarkets, fruit and vegetable markets, and natural food stores was associated with lower BMI	positive	New York NY	
13	restaurant, food_store	Association of food outlets and BMI varied depending on the adjustement for area or population, increased fruit and vegetable prices were associated with higher BMI, fast food prices were not associated with BMI	mixed	United States	

Table Id	Food Place Types	Summary of Results	Assessment of Results	Geographic Location	Location Notes
14	food_stores	Increase in energy dense food shelf space was associated with higher BMI	positive	Louisiana	southestern part of the state
15	fast_food, restaurant	Higher concentration of restaurants is associated with higher BMI, car owners have higher BMI than non-car owners, non-car owner in high fast food concentration have higher BMI than car owner in areas with no fast food	positive	Los Angeles CA	
16	fast_food, convenience	Presence of conveneince store was associated with BMI in top tertile.	positive	New York NY	East Harlem
17	convenience, restaurant, fast_food, grocery	Having one or more grocery options is assiciated with lower BMI in low income neighborhoods.	mixed	Salt Lake City UT	
18	supermarket, food_store	Neighborhood obesity was associated with availability of supermarkets and food stores.	positive	New York NY	
19	supermarket_chain, supermarket_independent, market_small, convenience	More chain supermarkets per roadway mile was associated with lower BMI for individuals without chronic conditions.	positive	Los Angeles CA	

Table Id	Food Place Types	Summary of Results	Assessment of Results	Geographic Location	Location Notes
20	supermarket, food_store	Neighborhood deprivation was associated with BMI, but the presence of supermarkets or other retail food stores did not mediate the relationship.	negative	Kansas	
21	fast_food, restaurant_chain, restaurant_independent	County wide BMI was positively associated with more fast food and chain restaurants and negatively associated with independent restaurants, poverty and sprawl were also associated with BMI.	positive	California	
22	supermarket, grocery, food_stores	As supermarket/grocery store availability increased from the 10th to 90th percentile, BMI decreased, as fast food availability increased (10th to 90th), BMI increased; neighborhood SES was associated with BMI, supermarket/grocery was associated with BP.	positive	United States	
23	fast_food	Proximity of fast food to home or work was not associated with BMI.	negative	Minnesota	
24	food_stores, fast_food, convenience, restaurant, food_other	Restaurants and other food places were associated with better diet, convenience stores were not associated with BMI (NS but trend in hypothesized direction).	mixed	Baltimore MD	

Table Id	Food Place Types	Summary of Results	Assessment of Results	Geographic Location	Location Notes
25	restaurant, fast_food, grocery, retail_other	BMI was positively associated with presence of convenience store within 1600m buffer, sugar sweetened bev intake was associated with restaurant, fast food, convenience store, grocery stores, and retail facilities in 1600m buffer around home; other diet measures were not associated with the residential environment.	mixed	Minneapolis MN	
26	fast_food, supermarket_chain	Compared two different data sources of food environment, and association with county level BMI; variation in agreement between two measures was ~50-100%; fast food per capital was associated with BMI, as was the retail food environment index.	mixed	North Carolina	
27	NEMS_score	In predominantly white neighborhoods there was a positive assocation between healthy food availability and BMI.	negative	Baltimore MD	
28	market_farmer, fast_food, pizza,	Network distance to convenience store was negatively associated with BMI, distance to farmers market was positively associated.	positive	Pitt County NC	
29	fast_food	No consistent association between fast food availability and BMI, some association in younger adults.	negative	Jackson MS	

Table Id	Food Place Types	Summary of Results	Assessment of Results	Geographic Location	Location Notes
30		Among some subgroups a broad measure of neighborhood amenities that included food-related locations was associated with physical activity but not BMI.	mixed	California	
31	supermarket_chain, supermarket_brand, supermarket_discount	BMI was higher in specific supermarket brands including discount stores.	positive	Paris France	
32	restaurant, convenience, supermarket, retail_other	Some factors were associated with BMI in girls; clusters of factors provided complex mixes of factors associated with BMI.	mixed	Minneapolis MN	

Table 30: Literature Review - Part 3 Studies - Citations and Measure Description

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
33	3	16140349	Sturm R, Datar A. Body mass index in elementary school children, metropolitan area food prices and food outlet density. Public Health 2005 Dec;119(12):1059-68.	132	longitudinal, bmi, individual, school, residential, children, prices	population adjusted density by ZIP	ZIP	school
34	3	19732982	Powell LM. Fast food costs and adolescent body mass index: evidence from panel data. J Health Econ 2009 Sep;28(5):963-70.	14	bmi, longitudinal, adolescent, price, individual	availability		
35	3	21665062	Leung CW, Laraia BA, Kelly M, Nickleach D, Adler NE, Kushi LH, Yen IH. The influence of neighborhood food stores on change in young girls' body mass index. Am J Prev Med 2011 Jul;41(1):43-51.	4	children, bmi, longitudinal	availability in buffer	network buffer of 0.25 and 1 mile	
36	3	21088263	Gibson DM. The neighborhood food environment and adult weight status: estimates from longitudinal data. Am J Public Health 2011 Jan;101(1):71-8.	3	adult, bmi, longitudinal	density (type?) in ZIP	ZIP	
37	3	21965186	Block JP, Christakis NA, O'Malley AJ, Subramanian SV. Proximity to food establishments and body mass index in the Framingham Heart Study offspring cohort over 30 years. Am J Epidemiol 2011 Nov 15;174(10):1108-14.	3	adult, bmi, framingham, longitudinal	network distance to nearest from home	[no zone]	

Table Id	Dissertation Part	Pubmed ID	Citation	Citation Count	Keywords	Imputed Measure	Measure Area	Measure Area (other)
38	3	22513496	Auchincloss AH, Mujahid MS, Shen M, Michos ED, Whitt-Glover MC, Diez Roux AV. Neighborhood health-promoting resources and obesity risk (the Multi-Ethnic Study of Atherosclerosis). Obesity (Silver Spring) 2012 Apr 19.		bmi, adult, MESA, longitudinal, obesity, individual	perception of healthy food in buffer	1 mile buffer around home	

Table 31: Literature Review - Part 3 Studies - Summary of Results

Table Id	Food Place Types	Summary of Results	Assessment of Results	Geographic Location	Location Notes
33	restaurant, grocery, convenience	No effect of outlet density on change in BMI, some effect of fruit and veggie prices on BMI, 4 years of follow-up.	negative	United States	
34	fast_food, restaurant, supermarket, grocery, convenience	Price of fast food but not availability is associated with BMI.	mixed	United States	
35	convenience, market_farmer,	More farmers marketsin 1 mile was associated with lower BMI, 3 year follow up, more convenience stores in 0.25 mile was associated with higher BMI.	positive		
36	grocery_small, supermarket, restaurant	In urban areas higher density of small grocery stores was associated with higher BMI.	positive		
37	supermarker, grocery, fast_food	For a 1 km increase in distance to nearest fast food, 0.11 unit decrease in BMI, other measures showed inconsistent results.	mixed	Framingham MA	Framingham cities in western suburbs of Boston
38	food_health_perception	Healthy food environment was associated with lower incidence of BMI as neighborhood score increased (but to null with full adjustment).	mixed	United States	6 regions

Appendix D – Size and Distance-based Store Probabilities

A continuous probability for shopping in a chain supermarket could be defined by:

a) P^i = retail sales (R) for the chain supermarket (sm) divided by the sum of retail sales for alternative food places (k), or $P^i = \frac{R_{sm}}{\sum R_k}$

b) P^i = sum retail sales (R) for the chain supermarket (sm) divided by the sum of retail sales for alternative food places (k), or $P^i = \frac{\sum R_{sm}}{\sum R_k}$

Adding distance to store enhances the probabilities estimated by retail sales (R) alone and better matches classical gravity models, which typically incorporate both distance and size (retail sales is a size proxy) to estimate probability of visiting a given location. For example, 3(a) would be modified to:

a) P^i = sum retail sales (R) for the chain supermarket (sm) divided by the sum of retail sales for alternative food places (k) multiplied by the inverse of the distance (d) for the chain supermarket (or k chain supermarkets) divided by the sum of the distance to all alternative food places (k), or $P^i = \frac{\sum R_{sm}}{\sum R_k} \times \sum \frac{1}{d_{sm}} \Big/ \sum \frac{1}{d_k}$

Appendix E – Detailed Model Results for Part 1

Table 32: Observed Chain Shopping Controlling for Imputed Chain Shopping, Individual and Neighborhood Characteristics

Chain Shopping Observed	Dep. Var.	Dep. Var.	Dep. Var.	Dep. Var.	Dep. Var.
Model Number	2013-1-17-7	2013-1-17-8	2013-1-17-9	2012-12-4-13	2013-1-16-3
Observations	2297	2297	2297	2297	2297
Parameter Count	6	8	10	17	16
Wald Chi Squared	193.44	197.75	224.94	316.18	193.25
Chi Squared Test P-Value	0	0	0	0	0
Log Likelihood	-995.05184	-991.01659	-990.46496	-985.724	-1045.7536
Pseudo R-squared	0.2575	0.2605	0.2609	0.2645	0.2197
Chain Shopping Imputed	6.55 (3.57-12.04)*	6.71 (3.62-12.42)*	6.79 (3.70-12.43)*	6.89 (3.80-12.50)*	
Age (years)				1.00 (0.99-1.02)	1.00 (0.99-1.01)
Female				1.05 (0.69-1.61)	1.02 (0.66-1.56)
Household with Children				0.91 (0.61-1.34)	0.88 (0.60-1.30)
Latino	0.41 (0.27-0.61)*	0.46 (0.31-0.70)*	0.42 (0.24-0.74)*	0.45 (0.25-0.82)*	0.50 (0.28-0.91)*
African American or Black			0.80 (0.36-1.78)	0.82 (0.38-1.76)	0.97 (0.45-2.09)
Other Race or Ethnic Group			0.87 (0.42-1.78)	0.92 (0.46-1.86)	0.94 (0.49-1.82)
Education - Less Than High School	0.45 (0.33-0.62)*	0.34 (0.17-0.66)*	0.34 (0.17-0.68)*	0.35 (0.17-0.71)*	0.42 (0.20-0.87)*
Education - High School		0.60 (0.29-1.24)	0.61 (0.28-1.31)	0.63 (0.30-1.36)	0.69 (0.32-1.47)
Education - Some College		0.84 (0.46-1.55)	0.87 (0.46-1.62)	0.91 (0.49-1.67)	0.90 (0.48-1.69)
Household Income (dollars)				1.00 (1.00-1.00)	1.00 (1.00-1.00)
Employed				0.99 (0.68-1.45)	0.99 (0.67-1.45)
Married or Living with Partner				0.95 (0.68-1.32)	0.99 (0.72-1.35)
US Born	2.28 (1.54-3.38)*	2.35 (1.60-3.45)*	2.31 (1.62-3.28)*	2.23 (1.57-3.15)*	2.17 (1.52-3.09)*
Household Owns Automobile				0.74 (0.51-1.07)	0.73 (0.49-1.07)
Poverty - Very Poor	0.22 (0.11-0.44)*	0.23 (0.11-0.49)*	0.25 (0.13-0.48)*	0.25 (0.12-0.49)*	0.15 (0.07-0.33)*
Poverty - Poor	0.51 (0.27-0.94)*	0.52 (0.28-0.96)*	0.53 (0.29-0.96)*	0.53 (0.29-0.98)*	0.42 (0.20-0.89)*

[Constant]	4.50 (2.66-7.62)*	5.25 (3.00-9.21)*	5.65 (2.94-10.88)*	5.91 (1.97-17.69)*	9.27 (2.80-30.68)*

* $p < 0.05$

Results are odds ratios and 95% confidence intervals. All models are survey weighted and account for clustering by census tract. Race and ethnicity categories are Latino, White, African American or Black, and Other category is primarily composed of Asian groups. Education categories are less than high school, high school, some college, and college or higher. Poverty categories are very poor, poor, and non-poor. When interpreting parameter estimates the referent category is the category not listed in the table within the same column and may be more than one category. In tables with multiple models the referent categories may differ between models.

Table 33: Observed Chain Shopping - Full Models Comparing the Association with Imputed Chain Shopping

Chain Shopping Observed	Dep. Var.	Dep. Var.	Dep. Var.	Dep. Var.
Model Number	2012-12-4-13	2013-1-16-3	2013-1-16-2	2012-12-6-10
Observations	2297	2297	2297	2297
Parameter Count	17	16	16	15
Wald Chi Squared	316.18	193.25	392.12	212.19
Chi Squared Test P-Value	0	0	0	0
Log Likelihood	-985.724	-1045.7536	-974.01368	-1061.5421
Pseudo R-squared	0.2645	0.2197	0.2732	0.2079
Chain Shopping Imputed	6.89 (3.80-12.50)*		9.19 (4.52-18.68)*	
Age (years)	1.00 (0.99-1.02)	1.00 (0.99-1.01)	1.00 (0.99-1.01)	1.00 (0.99-1.01)
Female	1.05 (0.69-1.61)	1.02 (0.66-1.56)	1.08 (0.71-1.63)	1.06 (0.71-1.59)
Household with Children	0.91 (0.61-1.34)	0.88 (0.60-1.30)	0.96 (0.65-1.43)	0.93 (0.64-1.35)
Latino	0.45 (0.25-0.82)*	0.50 (0.28-0.91)*	0.54 (0.29-1.01)	0.54 (0.28-1.02)
African American or Black	0.82 (0.38-1.76)	0.97 (0.45-2.09)	0.86 (0.39-1.88)	0.80 (0.36-1.79)
Other Race or Ethnic Group	0.92 (0.46-1.86)	0.94 (0.49-1.82)	1.09 (0.54-2.19)	1.10 (0.58-2.07)
Education - Less Than High School	0.35 (0.17-0.71)*	0.42 (0.20-0.87)*	0.49 (0.26-0.92)*	0.58 (0.29-1.17)
Education - High School	0.63 (0.30-1.36)	0.69 (0.32-1.47)	0.83 (0.40-1.71)	0.90 (0.44-1.86)
Education - Some College	0.91 (0.49-1.67)	0.90 (0.48-1.69)	1.12 (0.62-2.03)	1.13 (0.61-2.09)
Household Income (dollars)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
Employed	0.99 (0.68-1.45)	0.99 (0.67-1.45)	0.99 (0.67-1.45)	0.99 (0.68-1.43)
Married or Living with Partner	0.95 (0.68-1.32)	0.99 (0.72-1.35)	0.99 (0.71-1.36)	1.02 (0.76-1.37)
US Born	2.23 (1.57-3.15)*	2.17 (1.52-3.09)*	2.48 (1.84-3.35)*	2.46 (1.78-3.40)*
Household Owns Automobile	0.74 (0.51-1.07)	0.73 (0.49-1.07)	0.72 (0.49-1.07)	0.76 (0.52-1.13)
Poverty - Very Poor	0.25 (0.12-0.49)*	0.15 (0.07-0.33)*		
Poverty - Poor	0.53 (0.29-0.98)*	0.42 (0.20-0.89)*		
Neighborhood Disadvantage (Wink.-Cub.)			2.17 (1.49-3.15)*	2.19 (1.42-3.38)*
[Constant]	5.91 (1.97-17.69)*	9.27 (2.80-30.68)*	3.33 (1.16-9.61)*	4.64 (1.56-13.75)*

* $p < 0.05$

Results are odds ratios and 95% confidence intervals. All models are survey weighted and account for clustering by census tract. Race and ethnicity categories are Latino, white, African American or black, and other category is primarily composed of Asian groups. Education categories are less than high school, high school, some college, and college or higher. Poverty categories are very poor, poor, and non-poor. When interpreting parameter estimates the referent category is the category not listed in the table within the same column and may be more than one category. In tables with multiple models the referent categories may differ between models.

Table 34: Comparison of Individual Predictors of Imputed Chain Shopping Compared to Observed Chain Shopping With and Without Imputation Control

Chain Shopping Observed		Dep. Var.	Dep. Var.	Dep. Var.
Chain Shopping Imputed	Dep. Var.			
Model Number	2012-12-6-11	2012-12-6-10	2013-1-16-2	2012-12-4-13
Observations	2297	2297	2297	2297
Parameter Count	15	15	16	17
Wald Chi Squared	58.22	212.19	392.12	316.18
Chi Squared Test P-Value	0	0	0	0
Log Likelihood	-1168.0782	-1061.5421	-974.01368	-985.724
Pseudo R-squared	0.0285	0.2079	0.2732	0.2645
Chain Shopping Imputed			9.19 (4.52-18.68)*	6.89 (3.80-12.50)*
Age (years)	0.99 (0.98-1.00)*	1.00 (0.99-1.01)	1.00 (0.99-1.01)	1.00 (0.99-1.02)
Female	0.99 (0.79-1.26)	1.06 (0.71-1.59)	1.08 (0.71-1.63)	1.05 (0.69-1.61)
Household with Children	1.17 (0.74-1.84)	0.93 (0.64-1.35)	0.96 (0.65-1.43)	0.91 (0.61-1.34)
Latino	0.66 (0.26-1.62)	0.54 (0.28-1.02)	0.54 (0.29-1.01)	0.45 (0.25-0.82)*
African American or Black	0.67 (0.20-2.29)	0.80 (0.36-1.79)	0.86 (0.39-1.88)	0.82 (0.38-1.76)
Other Race or Ethnic Group	1.13 (0.44-2.95)	1.10 (0.58-2.07)	1.09 (0.54-2.19)	0.92 (0.46-1.86)
Education - Less Than High School	1.86 (0.70-4.90)	0.58 (0.29-1.17)	0.49 (0.26-0.92)*	0.35 (0.17-0.71)*
Education - High School	2.25 (0.95-5.30)	0.90 (0.44-1.86)	0.83 (0.40-1.71)	0.63 (0.30-1.36)
Education - Some College	1.47 (0.58-3.68)	1.13 (0.61-2.09)	1.12 (0.62-2.03)	0.91 (0.49-1.67)
Household Income (dollars)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
Employed	1.09 (0.77-1.54)	0.99 (0.68-1.43)	0.99 (0.67-1.45)	0.99 (0.68-1.45)
Married or Living with Partner	1.14 (0.80-1.60)	1.02 (0.76-1.37)	0.99 (0.71-1.36)	0.95 (0.68-1.32)
US Born	1.37 (0.69-2.75)	2.46 (1.78-3.40)*	2.48 (1.84-3.35)*	2.23 (1.57-3.15)*
Household Owns Automobile	1.22 (0.87-1.70)	0.76 (0.52-1.13)	0.72 (0.49-1.07)	0.74 (0.51-1.07)
Poverty - Very Poor				0.25 (0.12-0.49)*
Poverty - Poor				0.53 (0.29-0.98)*
Neighborhood Disadvantage (Wink.-Cub.)	0.83 (0.34-2.00)	2.19 (1.42-3.38)*	2.17 (1.49-3.15)*	
[Constant]	0.17 (0.03-0.85)*	4.64 (1.56-13.75)*	3.33 (1.16-9.61)*	5.91 (1.97-17.69)*

* $p < 0.05$

Results are odds ratios and 95% confidence intervals. All models are survey weighted and account for clustering by census tract. Race and ethnicity categories are Latino, white, African American or black, and other category is primarily composed of Asian groups. Education categories are less than high school, high school, some college, and college or higher. Poverty categories are very poor, poor, and non-poor. When interpreting parameter estimates the referent category is the category not listed in the table within the same column and may be more than one category. In tables with multiple models the referent categories may differ between models.

Table 35: Imputation Match with Chain Shopping (Positive or Negative Match) Controlling for Individual and Neighborhood Characteristics (Limited and Full Models)

Chain Shopping Match yes-no	Dep. Var.	Dep. Var.	Dep. Var.	Dep. Var.	Dep. Var.
Model Number	2012-12-4-22	2012-12-4-25	2012-12-4-26	2012-12-4-32	2013-1-17-10
Observations	2297	2297	2297	2297	2297
Parameter Count	3	8	9	16	15
Wald Chi Squared	22.98	68.77	69.34	86.97	90.59
Chi Squared Test P-Value	0	0	0	0	0
Log Likelihood	-1547.797	-1445.3447	-1436.2269	-1421.7924	-1383.6989
Pseudo R-squared	0.0425	0.1059	0.1115	0.1204	0.144
Age (years)		0.99 (0.98-1.00)	0.99 (0.98-1.00)	1.00 (0.99-1.01)	1.00 (0.99-1.01)
Female		0.86 (0.65-1.14)	0.85 (0.64-1.12)	0.87 (0.66-1.14)	0.88 (0.65-1.17)
Household with Children	1.80 (1.30-2.49)*	1.39 (0.95-2.03)	1.47 (1.01-2.14)*	1.37 (0.93-2.01)	1.31 (0.87-1.99)
Latino				1.42 (0.74-2.72)	1.05 (0.53-2.07)
African American or Black				1.09 (0.50-2.37)	0.78 (0.34-1.76)
Other Race or Ethnic Group				1.01 (0.51-2.02)	0.87 (0.43-1.75)
Education - Less Than High School		6.02 (2.94-12.32)*	5.08 (2.57-10.05)*	4.13 (2.21-7.74)*	2.91 (1.53-5.53)*
Education - High School		2.97 (1.59-5.55)*	2.65 (1.44-4.89)*	2.57 (1.39-4.74)*	1.96 (1.00-3.84)*
Education - Some College		1.65 (0.92-2.95)	1.50 (0.85-2.65)	1.52 (0.85-2.74)	1.27 (0.67-2.39)
Household Income (dollars)			1.00 (1.00-1.00)*	1.00 (1.00-1.00)*	1.00 (1.00-1.00)*
Employed				1.13 (0.78-1.63)	1.17 (0.82-1.67)
Married or Living with Partner				0.99 (0.72-1.37)	0.95 (0.67-1.34)
US Born				0.72 (0.47-1.10)	0.73 (0.45-1.19)
Household Owns Automobile				1.14 (0.81-1.61)	1.25 (0.90-1.74)
Poverty - Very Poor	2.90 (1.25-6.75)*	1.64 (0.68-3.94)	1.49 (0.62-3.61)	1.36 (0.54-3.39)	
Poverty - Poor	1.88 (0.73-4.81)	1.33 (0.49-3.61)	1.25 (0.46-3.43)	1.08 (0.38-3.07)	
Neighborhood Disadvantage (Wink.-Cub.)					0.54 (0.30-0.96)*
[Constant]	0.42 (0.20-0.90)*	0.36 (0.16-0.81)*	0.46 (0.20-1.09)	0.41 (0.13-1.29)	0.48 (0.16-1.50)

* $p < 0.05$

Results are odds ratios and 95% confidence intervals. All models are survey weighted and account for clustering by census tract. Race and ethnicity categories are Latino, white, African American or black, and other category is primarily composed of Asian groups. Education categories are less than high school, high school, some college, and college or higher. Poverty categories are very poor, poor, and non-poor. When interpreting parameter estimates the referent category is the category not listed in the table within the same column and may be more than one category. In tables with multiple models the referent categories may differ between models.

Appendix F – Rationale for Statistical Methods

Multilevel models have become standard practice in the public health literature (Kawachi and Berkman 2003).[74] Multilevel models are common because they have characteristics that are compatible with both the data structure and the research questions asked in the field. MLMs have the following important characteristics: 1) ability to model data with complex structures, notably data in a nested/hierarchical form, for example data observations nested in Census tracts; 2) MLMs explicitly model variance ("heterogeneity"), in that given a hierarchical unit with observations (e.g. individuals in a Census tract), the outcome for observations in a hierarchical unit varies from unit to unit; 3) MLMs model "dependency," in time, space, or other contexts, for example, that an outcome among individuals is similar within the same hierarchical unit; and 4) MLMs broadly assess "contextuality" or "micro and macro relations" by assessing how a individual outcome is influenced by both individual characteristics and hierarchical unit characteristics.

Store types are not a random classification (are a fixed classification) and hence are treated as a variable and not a level. Store assignment can be considered random from a larger population of stores and hence is a level of analysis.

[74] This discussion is based on a course attended by the author taught by Kelvyn Jones and S.V. Subramanian at the University of California Santa Barbara in 2011. For additional reference see "Multilevel methods for public health research" in Kawachi (2003).

There are alternatives to multilevel modeling. The first, which is dismissed given the desire to understand an outcome within individuals, is to take the means for a group and model the outcome based on these means. This clearly is inappropriate and is the origin of ecologic or aggregation fallacy.

The next alternative would be to use ordinary least squared ("OLS") regression with individual level data assuming that each observation is independent. Given the presumed dependency of individuals in nested/hierarchical structure leads to underestimation of standard errors and Type I errors (concluding there is difference when there is none).

Next, the modeling approach could include both individual and higher unit predictors. This assumes that all group-level variance can be explained by the group level predictors and gives incorrect standard errors for the group level predictors. This is often called a "contextual" analysis.

Another alternative is to include a dummy variable for each group in the higher level, called a "fixed effects model" or analysis of covariance. This approach creates some problems: 1) if the groups are very numerous (e.g. households), 2) there is no single parameter that assesses the differences between groups, 3) it is not possible to make inferences beyond the groups in the sample, 4) group-level predictors cannot be included because all degrees of freedom are consumed by the dummies, 5) the conceptual target of inference in fixed effects models are the individual group-level units compared to inference about the effect of units generally.

One alternative to address clustering and incorrect estimation of standard errors is to use generalized estimating equations (GEE) approach, which specifically adjusts standard errors for the effect of clustering. This approach is limited because it considers the nesting of data a "nuisance" and not a focus of inference, cannot assess the variance that exists between groups, and cannot be extended to more than two levels of hierarchy or other complex structures (like cross classification).

Hence the multi-level or "random effects" approach has the characteristics of: 1) partitioning variance between what exists at level 1 (individuals in a group) and in the level 2 unit (between the group), often called "within group" and "between group" variance components, 2) corrects standard errors, 3) allows for "un-observables" at each level (I believe these are commonly called "latent effects"), and 4) estimates "micro" (individual level) models and "macro" (higher level) models simultaneously.

In short, MLMs model means, intercepts, slopes and variances, and partition the variance between levels. "Random effects" means allowing the intercept or slopes to vary (i.e. be modeled). In the most basic form, this random effect can be "null" (no covariates) and in that case you are essentially just partitioning the variance.

OLS has the assumptions of IID, or errors, which are "independently and identically distributed" with a mean zero. This means the error has constant variability ("homoscedastic") and that the residuals have no pattern (are "independent"). Together, these assumptions are often noted: $[e_i] \sim N(0, \sigma_e^2)$

Example: Two-level random-intercept multilevel model

This example presents a two level random intercept model. In this case the two levels are composed of individuals (at level 1) and neighborhoods, or Census tracts (level 2) in which individuals are nested within tracts, but are only associated with a single tract. The random intercept model allows the intercept at level 1 to vary as a function of parameters in the level 2 model.[75] This example will include one covariate.

This example[76] will use body mass index and age[77] as the dependent and independent variables respectively. There are two models to specify, the level 1 model which models BMI as a function of age within the neighborhood, and the level 2 model which models the average BMI in a neighborhood as a function of the average BMI across all neighborhoods and the difference in BMI from this average for each neighborhood. So the level 1 model can be described:

---

[75] The modeling approach can be as simple as being "null" in that case level 2 unit exist and are allowed to vary, but there are no covariates in the level 1 or 2 model, and the result partitions the variance between levels.
[76] Adapted from an example with house prices as the dependent variable and number of rooms as the independent variable.
[77] The relationship between BMI and age may actually be "U" shaped, but for this example consider it linear in an age range from 20 to 50 years, with mean age of 35.

BMI of an individual = BMI of the average aged individual within a neighborhood + BMI difference associated with age in all neighborhoods + “differential” in BMI for each individual in the neighborhood, or generally,

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + e_{ij}$$

where *i* denotes the individuals, *j* the neighborhoods, *y* in this example is BMI, *x* is age, β_{0j} and β_1 are the intercept and slope of a linear regression line and *e* the error.

The level 2 model can be described as:

BMI of the average aged individual within a neighborhood = BMI of the averaged age individual across all neighborhoods + “differential” in BMI for an averaged aged individual for each neighborhood, or generally,

$$\beta_{0j} = \beta_0 + u_{0j}$$

where *u* represents the “differential” and β_0 is the mean BMI of average age in the entire sample.

These two models can be combined by algebraic substitution and rearranged to yield:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + (u_{0j} + e_{ij})$$

described as, BMI of an individual = BMI of the averaged age individual across all neighborhoods + BMI difference associated with age in all neighborhoods + “differential” in BMI for an averaged aged individual for each neighborhood + "differential" in BMI for each individual in the neighborhood.

The IID assumptions are noted:

$$[u_{0j}] \sim N(0, \sigma_{u0}^2)$$

$$[e_{ij}] \sim N(0, \sigma_e^2)$$

There is an additional assumption that neighborhood and individual differentials are independent. This is noted:

$$Cov[u_{0j}, e_{ij}] = 0$$

Since one aim of MLMs is to estimate the variance between levels, the σ_{u0}^2 and σ_e^2 have specific interpretations. In terms of the example, the first parameter ("sigma squared 'u' zero") gives the between neighborhood variance in BMI controlling for age. The second parameter ("sigma squared error") gives the within neighborhood, between individual variance in BMI controlling for age.

Example: Two-level random intercept model with predictors at both levels

Predictors can be introduced at both levels in the analysis. Level 1 predictors are individual characteristics. Level 2 predictors are commonly called environmental or ecologic measures. The form of the level 2 model with one predictor is:

$$\beta_{0j} = \beta_0 + \beta_2 x_{2j} + u_{0j}$$

Substituting that into the unchanged level 1 model gives:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2j} + (u_{0j} + e_{ij})$$

Appendix G – Detailed Model Results for Part 2

Table 36: Main Analysis Model Results BMI and Store Type, BMI Without Store Type, Wave 1 Comparison

	Main (n = 902)	No store (n = 902)	Main W1 (n = 2294)
R-squared	0.1429	0.1224	0.1025
RMSE	5.5851	5.6289	4.6684
	Beta (LCI – UCI)	Beta (LCI – UCI)	Beta (LCI – UCI)
Poverty category (ref. Non-poor)			
Very poor	2.46 (0.58-4.35)*	2.17 (0.31-4.04)*	0.61 (-0.47-1.68)
Poor	1.69 (0.16-3.23)*	1.53 (0.15-2.91)*	0.84 (-0.06-1.74)
Spanish-language interview	0.55 (-1.42-2.52)	0.16 (-1.78-2.11)	
Own car	-0.75 (-2.10-0.60)	-0.74 (-2.05-0.57)	
Family income (logged in W2, not in W1)	0.14 (-0.30-0.59)	0.09 (-0.40-0.59)	0.00 (0.00-0.00)
Family income imputed	-0.94 (-1.94-0.07)	-1.05 (-2.10-0.00)	-1.10 (-1.72--0.47)*
Never exercise	1.42 (-0.30-3.13)	1.31 (-0.44-3.07)	
Never eat fast food	-1.01 (-2.74-0.73)	-1.04 (-2.78-0.69)	
Employed	-1.13 (-2.51-0.26)	-0.97 (-2.38-0.44)	0.20 (-0.63-1.03)
Married/Partner	-0.60 (-1.83-0.63)	-0.71 (-1.98-0.56)	0.62 (-0.05-1.29)
Family size	0.51 (0.07-0.96)*	0.53 (0.07-0.99)*	0.15 (-0.05-0.36)
Distance to store	0.02 (-0.08-0.12)	-0.01 (-0.09-0.07)	
Household with children	0.73 (-1.16-2.62)	0.81 (-0.96-2.59)	0.77 (0.07-1.47)*
Store type (ref Major - higher poverty)			
Discount	-1.81 (-4.17-0.55)		-0.43 (-1.52-0.65)
Specialty	-2.81 (-5.03--0.58)*		-2.31 (-3.94--0.69)*
Spanish-language	-1.96 (-3.71--0.21)*		-1.01 (-2.05-0.03)
Independent	-0.74 (-4.54-3.05)		-0.62 (-1.95-0.71)
Bulk	-2.23 (-5.42-0.96)		0.86 (-1.28-2.99)
Small	-2.18 (-5.36-1.00)		-1.19 (-2.50-0.13)
Major (lower poverty)	-0.87 (-3.05-1.32)		-0.51 (-1.86-0.84)
Age	0.01 (-0.04-0.05)	0.01 (-0.03-0.06)	0.03 (0.01-0.06)*
Female	-0.29 (-1.49-0.91)	-0.33 (-1.58-0.92)	-1.14 (-1.77--0.52)*
Latino	1.24 (-0.65-3.12)	1.39 (-0.50-3.27)	1.23 (0.00-2.46)
Black	-0.20 (-2.73-2.34)	0.10 (-2.19-2.39)	0.48 (-0.49-1.45)
Other	-0.15 (-2.89-2.59)	0.06 (-2.58-2.69)	-0.57 (-1.93-0.79)
U.S. Born	1.33 (-0.53-3.20)	1.75 (-0.23-3.72)	1.01 (0.14-1.88)*
Smoke	-1.94 (-3.36--0.52)*	-2.01 (-3.61--0.41)*	-0.85 (-1.83-0.12)
Education (less than HS)	0.32 (-2.08-2.72)	0.49 (-1.89-2.86)	0.01 (-1.39-1.40)
Education (HS)	-0.04 (-2.44-2.36)	0.02 (-2.41-2.45)	0.40 (-0.93-1.72)
Education (some college)	-0.58 (-2.34-1.18)	-0.44 (-2.26-1.38)	0.43 (-0.60-1.46)
Constant	26.97 (21.24-32.70)*	25.77 (19.46-32.08)*	23.63 (21.04-26.21)*

* $p < 0.05$

The global F-test of the store type variable was $p = 0.2815$ in the wave 2 main analysis and $p = 0.0056$ in the wave 1 main analysis. LCI - lower 95% confidence interval. UCI – upper 95% confidence interval.

Table 37: Main Model Results Stratified by Latino or White

	Latino (n = 523)	White (n = 227)
R-squared	0.1099	0.3177
RMSE	5.8374	4.5367
	Beta (LCI – UCI)	Beta (LCI – UCI)
Poverty category (ref. Non-poor)		
Very poor	1.04 (-1.71-3.78)	2.48 (-4.32-9.28)
Poor	1.39 (-0.98-3.77)	0.70 (-1.68-3.08)
Spanish-language interview	-0.26 (-2.31-1.79)	
Own car	0.02 (-1.36-1.40)	-0.19 (-2.05-1.66)
Family income (logged)	-0.01 (-0.39-0.37)	0.14 (-0.49-0.77)
Family income imputed	-1.28 (-2.87-0.30)	-0.67 (-2.37-1.02)
Never exercise	1.77 (-0.94-4.47)	3.45 (0.23-6.66)*
Never eat fast food	-0.64 (-3.05-1.77)	0.24 (-1.73-2.21)
Employed	0.19 (-1.94-2.32)	-0.44 (-2.05-1.17)
Married/Partner	-0.58 (-2.29-1.13)	-1.24 (-3.23-0.75)
Family size	0.45 (-0.20-1.11)	0.20 (-0.53-0.94)
Distance to store	-0.06 (-0.13-0.02)	0.31 (0.03-0.59)*
Household with children	0.80 (-1.03-2.63)	2.07 (-0.45-4.60)
Store type	(ref Major - all pov.)	(ref Major - high pov.)
Discount	-1.30 (-3.42-0.82)	-2.65 (-6.76-1.47)
Specialty	-2.21 (-5.05-0.64)	-3.88 (-6.00--1.76)*
Spanish-language	-0.81 (-2.50-0.87)	-1.62 (-5.03-1.78)
Independent	1.66 (-2.75-6.06)	-6.22 (-9.42--3.03)*
Bulk	0.23 (-2.79-3.25)	-7.69 (-13.83--1.55)*
Small	-0.09 (-2.76-2.58)	-0.44 (-4.26-3.38)
Major (low poverty)		-1.54 (-3.46-0.38)
Age	0.02 (-0.05-0.09)	-0.01 (-0.08-0.05)
Female	-0.53 (-2.36-1.30)	-0.22 (-1.80-1.35)
U.S. Born	1.85 (-0.21-3.90)	-2.19 (-6.22-1.83)
Smoke	-0.89 (-2.80-1.01)	-3.31 (-6.24--0.38)*
Education (less than HS)	2.62 (-0.65-5.89)	-2.11 (-5.59-1.37)
Education (HS)	1.79 (-1.60-5.19)	2.68 (-1.27-6.63)
Education (some college)	-0.52 (-3.71-2.67)	0.92 (-0.68-2.53)
Constant	26.00 (17.52-34.49)*	29.23 (21.87-36.58)*

* $p < 0.05$
LCI - lower 95% confidence interval. UCI – upper 95% confidence interval.

Table 38: Main Model Results with Fruit and Vegetable Consumption

	Beta (LCI – UCI)
Poverty category (ref. Non-poor)	
Very poor	2.39 (0.60-4.19)*
Poor	1.66 (0.18-3.13)*
Report eating four or more fruit/veggies	-1.26 (-2.33--0.18)*
Spanish-language interview	0.57 (-1.43-2.57)
Own car	-0.69 (-2.03-0.65)
Family income (logged)	0.15 (-0.27-0.58)
Family income imputed	-0.86 (-1.83-0.11)
Never exercise	1.40 (-0.30-3.09)
Never eat fast food	-0.90 (-2.60-0.81)
Employed	-1.16 (-2.54-0.22)
Married/Partner	-0.55 (-1.80-0.70)
Family size	0.52 (0.06-0.97)*
Distance to store	0.00 (-0.09-0.10)
Household with children	0.76 (-1.13-2.65)
Store type (ref Major - higher pov.)	
Discount	-1.74 (-4.02-0.54)
Specialty	-2.72 (-4.92--0.52)*
Spanish-language	-1.75 (-3.49--0.02)*
Independent	-0.68 (-4.45-3.10)
Bulk	-2.21 (-5.44-1.02)
Small	-2.09 (-5.35-1.17)
Major (lower poverty)	-0.73 (-2.90-1.45)
Age	0.01 (-0.04-0.05)
Female	-0.14 (-1.40-1.11)
Latino	1.05 (-0.78-2.89)
Black	-0.23 (-2.66-2.20)
Other	-0.25 (-2.88-2.38)
U.S. Born	1.18 (-0.65-3.01)
Smoke	-1.98 (-3.45--0.51)*
Education (less than HS)	0.10 (-2.32-2.51)
Education (HS)	-0.25 (-2.74-2.24)
Education (some college)	-0.68 (-2.39-1.02)
Constant	27.37 (22.10-32.65)*

* $p < 0.05$, n = 899, R-squared = 0.1534, RMSE = 5.558
LCI - lower 95% confidence interval. UCI – upper 95% confidence interval.

Appendix H – Detailed Model Results for Part 3

Table 39: Model Results Store Change Model

	OR (LCI – UCI)
Store opening or closing	5.64 (3.34-9.53)*
Age	0.98 (0.96-0.99)*
Employed at both waves	0.91 (0.60-1.37)
Female	0.94 (0.59-1.52)
Income changed	1.00 (1.00-1.00)*
Income imputed	0.79 (0.40-1.57)
Smoked at both waves	1.44 (0.63-3.32)
Married/Partner at both waves	0.94 (0.62-1.42)
Family size change	0.91 (0.74-1.12)
U.S. born	0.26 (0.14-0.47)*
Own car at both waves	1.25 (0.79-1.99)
Very Poor	1.29 (0.54-3.06)
Poor	1.39 (0.78-2.46)
Household with children	0.71 (0.38-1.33)
Education (less than HS)	0.50 (0.20-1.20)
Education (HS)	0.61 (0.26-1.43)
Education (some college)	0.57 (0.31-1.02)
Latino	0.84 (0.38-1.83)
Black	2.09 (0.78-5.59)
Other	0.80 (0.30-2.14)
Distance to store (farther)	1.07 (0.96-1.18)
Population change	1.00 (0.99-1.00)
Total store change	1.25 (0.76-2.05)
Store type (ref Major - all pov.)	
Discount	2.22 (0.78-6.35)
Specialty	19.23 (8.52-43.42)*
Spanish-language	1.96 (0.71-5.36)
Independent	7.68 (2.65-22.23)*
Bulk	14.18 (4.25-47.32)*
Small	20.00 (3.29-121.62)*
Constant	0.90 (0.26-3.15)

* $p < 0.05$, n = 582, pseudo-R-squared = 0.2943, Log pseudolikelihood = -270.9
OR – Odds Ratio. LCI - lower 95% confidence interval. UCI – upper 95% confidence interval.

References

Alonso, William. 1964. *Location and Land Use; Toward a General Theory of Land Rent*. Cambridge: Harvard University Press.

Anas, Alex, Richard Arnott, and Kenneth Small. 1998. "Urban Spatial Structure." *Journal of Economic Literature* 36 (3) (September 1): 1426–1464.

Auchincloss, Amy H, Mahasin S Mujahid, Mingwu Shen, Erin D Michos, Melicia C Whitt-Glover, and Ana V Diez Roux. 2012. "Neighborhood Health-Promoting Resources and Obesity Risk (the Multi-Ethnic Study of Atherosclerosis).." *Obesity (Silver Spring)* (April 19). doi:10.1038/oby.2012.91.

Auchincloss, Amy, Rick Riolo, Daniel Brown, Jeremy Cook, and Ana Diez Roux. 2011. "An Agent-Based Model of Income Inequalities in Diet in the Context of Residential Segregation." *American Journal of Preventive Medicine* 40 (3): 303–311. doi:doi: DOI: 10.1016/j.amepre.2010.10.033.

Black, Jennifer L, James Macinko, L Beth Dixon, and George E Fryer. 2010. "Neighborhoods and Obesity in New York City.." *Health & Place* 16 (3) (May): 489–499. doi:10.1016/j.healthplace.2009.12.007.

Block, Jason P, Nicholas A Christakis, A James O'Malley, and S V Subramanian. 2011. "Proximity to Food Establishments and Body Mass Index in the Framingham Heart Study Offspring Cohort Over 30 Years.." *American Journal of Epidemiology* 174 (10) (November 15): 1108–1114. doi:10.1093/aje/kwr244.

Block, Jason P, Richard A Scribner, and Karen B DeSalvo. 2004. "Fast Food, Race/Ethnicity, and Income: a Geographic Analysis.." *American Journal of Preventive Medicine* 27 (3) (October): 211–217. doi:10.1016/j.amepre.2004.06.007.

Bourdieu, Pierre. 1984. *Distinction: a Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Breslow, L. 1996. "Social Ecological Strategies for Promoting Healthy Lifestyles." *American Journal of Health Promotion : AJHP* 10 (4): 253–257.

Brown, Arleen F, Roberto B Vargas, Alfonso Ang, and Anne R Pebley. 2008. "The Neighborhood Food Resource Environment and the Health of Residents with Chronic Conditions: the Food Resource Environment and the Health of Residents.." *Journal of General Internal Medicine* 23 (8) (August): 1137–1144. doi:10.1007/s11606-008-0601-5.

Casagrande, Sarah Stark, Manuel Franco, Joel Gittelsohn, Alan B Zonderman, Michele K Evans, Marie Fanelli Kuczmarski, and Tiffany L Gary-Webb. 2011. “Healthy Food Availability and the Association with BMI in Baltimore, Maryland..” *Public Health Nutr* 14 (6) (June): 1001–1007. doi:10.1017/S1368980010003812.

Chaix, Basile, Kathy Bean, Mark Daniel, Shannon N Zenk, Yan Kestens, Hélène Charreire, Cinira Leal, et al. 2012. “Associations of Supermarket Characteristics with Weight Status and Body Fat: a Multilevel Analysis of Individuals Within Supermarkets (RECORD Study)..” *PloS One* 7 (4): e32908. doi:10.1371/journal.pone.0032908.

Clifton, Kelly J. 2004. “Mobility Strategies and Food Shopping for Low-Income Families: a Case Study.” *Journal of Planning Education and Research* 23 (4) (June 1): 402–413. doi:10.1177/0739456X04264919.

Connors, M, Carole A Bisogni, Jeffery Sobal, and C M Devine. 2001. “Managing Values in Personal Food Systems.” *Appetite* 36 (3) (June 1): 189–200. doi:10.1006/appe.2001.0400.

Drewnowski, Adam, Anju Aggarwal, Philip M Hurvitz, Pablo Monsivais, and Anne V Moudon. 2012. “Obesity and Supermarket Access: Proximity or Price?.” *American Journal of Public Health* 102 (8) (August): e74–80. doi:10.2105/AJPH.2012.300660.

Dubowitz, Tamara, Madhumita Ghosh-Dastidar, Christine Eibner, Mary E Slaughter, Meenakshi Fernandes, Eric A Whitsel, Chloe E Bird, et al. 2011. “The Women's Health Initiative: the Food Environment, Neighborhood Socioeconomic Status, BMI, and Blood Pressure..” *Obesity (Silver Spring)* (June 9). doi:10.1038/oby.2011.141.

Escarce, José J, Nicole Lurie, and Adria Jewell. 2011. “RAND Center for Population Health and Health Disparities (CPHHD) Data Core Series: Decennial Census Abridged, 1990-2010 [United States].” ISCPR27866-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Ford, Paula B, and David A Dzewaltowski. 2011. “Neighborhood Deprivation, Supermarket Availability, and BMI in Low-Income Women: a Multilevel Analysis..” *Journal of Community Health* (March 10). doi:10.1007/s10900-011-9377-3.

Friedmann, J. 1987. *Planning in the Public Domain: From Knowledge to Action*. Princeton, N.J. : Princeton University Press.

Frumkin, Howard, Lawrence D Frank, and Richard Jackson. 2004. *Urban Sprawl and Public Health : Designing, Planning, and Building for Healthy Communities*.

Furst, T, M Connors, Carole A Bisogni, Jeffery Sobal, and Laura Winter Falk. 1996. “Food Choice: a Conceptual Model of the Process.” *Appetite* 26 (3) (June 1): 247–265. doi:10.1006/appe.1996.0019.

Galvez, Maida P, Lu Hong, Elizabeth Choi, Laura Liao, James Godbold, and Barbara Brenner. 2009. “Childhood Obesity and Neighborhood Food-Store Availability in an Inner-City Community..” *Academic Pediatrics* 9 (5) (September): 339–343. doi:10.1016/j.acap.2009.05.003.

Gibson, Diane M. 2011. “The Neighborhood Food Environment and Adult Weight Status: Estimates From Longitudinal Data..” *American Journal of Public Health* 101 (1) (January): 71–78. doi:10.2105/AJPH.2009.187567.

Glaeser, E, M Kahn, and J Rappaport. 2008. “Why Do the Poor Live in Cities the Role of Public Transportation.” *Journal of Urban Economics* 63 (1): 1–24. doi:10.1016/j.jue.2006.12.004.

Gregson, Jennifer. 2011. “Poverty, Sprawl, and Restaurant Types Influence Body Mass Index of Residents in California Counties..” *Public Health Reports (Washington, DC : 1974)* 126 Suppl 1 (May): 141–149.

Guthman, Julie. 2003. “Fast Food/Organic Food: Reflexive Tastes and the Making of ‘Yuppie Chow’.” *Social & Cultural Geography* 4 (1): 45–58.

Hall, Peter. 2002. *Cities of Tomorrow : an Intellectual History of Urban Planning and Design in the Twentieth Century*. Oxford, UK ; Malden, MA : Blackwell Publishers.

Hanson, S. 1995. *The geography of urban transportation*. New York : Guilford Press.

Handy, Susan. 1996. “Understanding the Link Between Urban Form and Nonwork Travel Behavior.” *Journal of Planning Education and Research* 15 (3) (April 1): 183–198.

Hickson, DeMarc, Ana Diez Roux, Adam Smith, Katherine Tucker, Larry Gore, Lei Zhang, and Sharon Wyatt. 2011. “Associations of Fast Food Restaurant Availability with Dietary Intake and Weight Among African Americans in the Jackson Heart Study, 2000-2004.” *American Journal of Public Health* (May

9): AJPH.2010.300006v2. doi:10.2105/AJPH.2010.300006.

Hillier, Amy, Carolyn C. Cannuscio, Allison Karpyn, Jacqueline McLaughlin, Mariana Chilton, and Karen Glanz. 2011. “How Far Do Low-Income Parents Travel to Shop for Food? Empirical Evidence From Two Urban Neighborhoods.” *Urban Geography* 32 (5): 712–729. doi:10.2747/0272-3638.32.5.712.

Horowitz, Carol R, Kathryn A Colson, Paul L Hebert, and Kristie J Lancaster. 2004. “Barriers to Buying Healthy Foods for People with Diabetes: Evidence of Environmental Disparities.” *American Journal of Public Health* 94 (9) (September 1): 1549–1554.

Huff, David L. 1963. “A Probabalistic Analysis of Shopping Center Trading Areas.” *Land Economics* 39 (1): 81–90.

Huxley, Margo, and Oren Yiftachel. 2000. “New Paradigm or Old Myopia? Unsettling the Communicative Turn in Planning Theory.” *Journal of Planning Education and Research* 19 (4) (June 1): 333. doi:10.1177/0739456X0001900402.

Inagami, S, Deborah A Cohen, Brian K Finch, and Steven M Asch. 2006. “You Are Where You Shop Grocery Store Locations, Weight, and Neighborhoods.” *American Journal of Preventive Medicine* 31 (1) (July 1): 10–17. doi:10.1016/j.amepre.2006.03.019.

Inagami, Sanae, Deborah A Cohen, Arleen F Brown, and Steven M Asch. 2009. “Body Mass Index, Neighborhood Fast Food and Restaurant Concentration, and Car Ownership..” *J Urban Health* 86 (5) (September): 683–695. doi:10.1007/s11524-009-9379-y.

Jeffery, Robert W, Judy Baxter, Maureen McGuire, and Jennifer Linde. 2006. “Are Fast Food Restaurants an Environmental Risk Factor for Obesity?.” *The International Journal of Behavioral Nutrition and Physical Activity* 3: 2. doi:10.1186/1479-5868-3-2.

Jilcott, Stephanie B, Jared T McGuirt, Satomi Imai, and Kelly R Evenson. 2010. “Measuring the Retail Food Environment in Rural and Urban North Carolina Counties..” *Journal of Public Health Management and Practice : JPHMP* 16 (5) (September): 432–440. doi:10.1097/PHH.0b013e3181bdebe4.

Jilcott, Stephanie B, Scott Wade, Jared T McGuirt, Qiang Wu, Suzanne Lazorick, and Justin B Moore. 2011. “The Association Between the Food Environment and Weight Status Among Eastern North Carolina Youth..” *Public Health Nutr*

14 (9) (September): 1610–1617. doi:10.1017/S1368980011000668.

Kain, John. 1968. “Housing Segregation, Negro Employment, and Metropolitan Decentralization.” *Quarterly Journal of Economics* 82 (2): 175–197.

Kawachi, Ichiro, and Lisa F Berkman. 2003. *Neighborhoods and Health*. Oxford ; New York : Oxford University Press.

Keegan, Theresa H M, Susan Hurley, Debbie Goldberg, David O Nelson, Peggy Reynolds, Leslie Bernstein, Pam L Horn-Ross, and Scarlett L Gomez. 2012. “The Association Between Neighborhood Characteristics and Body Size and Physical Activity in the California Teachers Study Cohort..” *American Journal of Public Health* 102 (4) (April): 689–697. doi:10.2105/AJPH.2011.300150.

Krieger, Nancy. 2011. *Epidemiology and the People's Health : Theory and Context*. New York : Oxford University Press.

Krugman, Paul R, and Robin Wells. 2009. *Microeconomics*. New York, NY : Worth Publishers.

Laska, Melissa N, Mary O Hearst, Ann Forsyth, Keryn E Pasch, and Leslie Lytle. 2010. “Neighbourhood Food Environments: Are They Associated with Adolescent Dietary Intake, Food Purchases and Weight Status?.” *Public Health Nutr* 13 (11) (November): 1757–1763. doi:10.1017/S1368980010001564.

Leung, Cindy W, Barbara A Laraia, Maggi Kelly, Dana Nickleach, Nancy E Adler, Lawrence H Kushi, and Irene H Yen. 2011. “The Influence of Neighborhood Food Stores on Change in Young Girls' Body Mass Index..” *American Journal of Preventive Medicine* 41 (1) (July): 43–51. doi:10.1016/j.amepre.2011.03.013.

Loong, Tze-Wey. 2003. “Understanding Sensitivity and Specificity with the Right Side of the Brain.” *Bmj* 327 (January 25).

Lopez, Russ P. 2007. “Neighborhood Risk Factors for Obesity..” *Obesity (Silver Spring)* 15 (8) (August): 2111–2119. doi:10.1038/oby.2007.251.

Mantovani, Richard E, and James Welsh. 1996. “Authorized Food Retailer Characteristics Study.” United States Department of Agriculture.

Massey, Douglas, and Nancy Denton. 1993. *American Apartheid : Segregation and the Making of the Underclass*.

Millstein, Rachel A, Hsin-Chieh Yeh, Frederick L Brancati, Marian Batts-Turner, and Tiffany L Gary. 2009. "Food Availability, Neighborhood Socioeconomic Status, and Dietary Patterns Among Blacks with Type 2 Diabetes Mellitus.." *Medscape Journal of Medicine* 11 (1): 15.

Mokdad, Ali H, James S Marks, Donna F Stroup, and Julie L Gerberding. 2004. "Actual Causes of Death in the United States, 2000.." *JAMA : the Journal of the American Medical Association* 291 (10) (March 10): 1238–1245. doi:10.1001/jama.291.10.1238.

Moore, Latetia V, and Ana V Diez Roux. 2006. "Associations of Neighborhood Characteristics with the Location and Type of Food Stores." *American Journal of Public Health* 96 (2) (February 1): 325–331. doi:10.2105/AJPH.2004.058040.

Morland, Kimberly B, Ana V Diez Roux, and S Wing. 2006. "Supermarkets, Other Food Stores, and Obesity: the Atherosclerosis Risk in Communities Study." *American Journal of Preventive Medicine* 30 (4) (April 1): 333–339. doi: 10.1016/j.amepre.2005.11.003 [doi].

Morland, Kimberly B, S Wing, Ana V Diez Roux, and C Poole. 2002. "Neighborhood Characteristics Associated with the Location of Food Stores and Food Service Places." *American Journal of Preventive Medicine* 22 (1): 23–29. doi:S0749379701004032 [pii].

Muth, Richard. 1969. *Cities and Housing; the Spatial Pattern of Urban Residential Land Use*. Chicago: University of Chicago Press.

Ohls, James C, Michael Ponza, Lorenzo Moreno, Amy Zambrowski, and Rhoda Cohen. 1999. "Food Stamp Participants Access to Food Retailers." United States Department of Agriculture.

Ong, Paul, and Douglas Miller. 2005. "Spatial and Transportation Mismatch in Los Angeles." *Journal of Planning Education and Research* 25 (1) (September 1): 43. doi:10.1177/0739456X04270244.

Powell, Lisa M. 2009. "Fast Food Costs and Adolescent Body Mass Index: Evidence From Panel Data.." *J Health Econ* 28 (5) (September): 963–970. doi:10.1016/j.jhealeco.2009.06.009.

Powell, Lisa M, and Yanjun Bao. 2009. "Food Prices, Access to Food Outlets and Child Weight.." *Econ Hum Biol* 7 (1) (March): 64–72. doi:10.1016/j.ehb.2009.01.004.

Powell, Lisa M, M Christopher Auld, Frank J Chaloupka, Patrick M O'Malley, and Lloyd D Johnston. 2007. “Associations Between Access to Food Stores and Adolescent Body Mass Index..” *American Journal of Preventive Medicine* 33 (4 Suppl) (October): S301–7. doi:10.1016/j.amepre.2007.07.007.

Powell, Lisa M, Sandy Slater, Donka Mirtcheva, Yanjun Bao, and Frank J Chaloupka. 2007. “Food Store Availability and Neighborhood Characteristics in the United States.” *Preventive Medicine* 44 (3) (March 1): 189–195. doi:10.1016/j.ypmed.2006.08.008.

Ritzer, G. 1993. *The McDonaldization of Society: an Investigation Into the Changing Character of Social Life*. Newbury Park, CA: Pine Forge Press.

Rose, Donald, Paul L Hutchinson, J Nicholas Bodor, Chris M Swalm, Thomas A Farley, Deborah A Cohen, and Janet C Rice. 2009. “Neighborhood Food Environments and Body Mass Index: the Importance of in-Store Contents..” *American Journal of Preventive Medicine* 37 (3) (September): 214–219. doi:10.1016/j.amepre.2009.04.024.

Rundle, Andrew, Kathryn M Neckerman, Lance Freeman, Gina S Lovasi, Marnie Purciel, James Quinn, Catherine Richards, Neelanjan Sircar, and Christopher Weiss. 2009. “Neighborhood Food Environment and Walkability Predict Obesity in New York City..” *Environmental Health Perspectives* 117 (3) (March): 442–447. doi:10.1289/ehp.11590.

Sallis, James F, P R Nader, J W Rupp, C J Atkins, and W C Wilson. 1986. “San Diego Surveyed for Heart-Healthy Foods and Exercise Facilities.” *Public Health Reports (Washington, DC : 1974)* 101 (2): 216–219.

Sastry, Narayan, Bonnie Ghosh-Dastidar, John Adams, and Anne R Pebley. 2006. “The Design of a Multilevel Survey of Children, Families, and Communities: the Los Angeles Family and Neighborhood Survey.” *Social Science Research* 35 (4) (December): 1000–1024. doi:10.1016/j.ssresearch.2005.08.002.

Schelling, Thomas. 1971. “Dynamic Models of Segregation.” *Journal of Mathematical Sociology* 1: 143–186.

Sloane, D, A Diamant, L Lewis, Antronette K Yancey, G Flynn, L Nascimento, W McCarthy, J Guinyard, and M Cousineau. 2003. “Improving the Nutritional Resource Environment for Healthy Living Through Community-Based Participatory Research.” *Journal of General Internal Medicine* 18 (7) (July 1): 568–575. doi:jgi21022 [pii].

Spiegelman, Donna. 2010. “Approaches to Uncertainty in Exposure Assessment in Environmental Epidemiology..” *Annual Review of Public Health* 31: 149–163. doi:10.1146/annurev.publhealth.012809.103720.

Stouffer, Samuel A. 1940. “Intervening Opportunities - a Theory Relating Mobility and Distance.” *American Sociological Review* 6 (6): 845–867.

Sturm, R, and A Datar. 2005. “Body Mass Index in Elementary School Children, Metropolitan Area Food Prices and Food Outlet Density..” *Public Health* 119 (12) (December): 1059–1068. doi:10.1016/j.puhe.2005.05.007.

Tiebout, Charles. 1956. “A Pure Theory of Local Expenditures.” *Journal of Political Economy* 64 (5): 416–424.

Wacholder, S. 1995. “When Measurement Errors Correlate with Truth: Surprising Effects of Nondifferential Misclassification..” *Epidemiology (Cambridge, Mass)* 6 (2) (March): 157–161.

Wachs, Martin. 1993. “Learning From Los Angeles: Transport, Urban Form, and Air Quality.” *Transportation* 20 (4): 329–354. doi:10.1007/BF01100463.

Wall, Melanie M, Nicole I Larson, Ann Forsyth, David C Van Riper, Dan J Graham, Mary T Story, and Dianne Neumark-Sztainer. 2012. “Patterns of Obesogenic Neighborhood Features and Adolescent Weight: a Comparison of Statistical Approaches..” *American Journal of Preventive Medicine* 42 (5) (May): e65–75. doi:10.1016/j.amepre.2012.02.009.

Wang, May C, Soowon Kim, Alma A Gonzalez, Kara E MacLeod, and Marilyn A Winkleby. 2007. “Socioeconomic and Food-Related Physical Characteristics of the Neighbourhood Environment Are Associated with Body Mass Index..” *Journal of Epidemiology and Community Health* 61 (6) (June): 491–498. doi:10.1136/jech.2006.051680.

Wang, Y, and M Beydoun. 2007. “The Obesity Epidemic in the United States--Gender, Age, Socioeconomic, Racial/Ethnic, and Geographic Characteristics: a Systematic Review and Meta-Regression Analysis.” *Epidemiol Rev* 29: 6–28. doi: 10.1093/epirev/mxm007.

Warde, Alan. 1997. *Consumption, Food, and Taste: Culinary Antinomies and Commodity Culture*. London: Sage Publications.

Wilson, WJ. 1987. “The Truly Disadvantaged - the Inner City, the Underclass, and Public Policy.”

Winkleby, M A, and C Cubbin. 2003. “Influence of Individual and Neighbourhood Socioeconomic Status on Mortality Among Black, Mexican-American, and White Women and Men in the United States..” *Journal of Epidemiology and Community Health* 57 (6) (June): 444–452.

Yach, Derek, Corinna Hawkes, C Linn Gould, and Karen J Hofman. 2004. “The Global Burden of Chronic Diseases: Overcoming Impediments to Prevention and Control..” *JAMA : the Journal of the American Medical Association* 291 (21) (June 2): 2616–2622. doi:10.1001/jama.291.21.2616.

Zenk, Shannon N, Amy J Schulz, Barbara A Israel, Sherman A James, Shuming Bao, and Mark L Wilson. 2005. “Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit..” *American Journal of Public Health* 95 (4) (April): 660–667. doi:10.2105/AJPH.2004.042150.

Zick, Cathleen D, Ken R Smith, Jessie X Fan, Barbara B Brown, Ikuho Yamada, and Lori Kowaleski-Jones. 2009. “Running to the Store? the Relationship Between Neighborhood Environments and the Risk of Obesity..” *Soc Sci Med* 69 (10) (November): 1493–1500. doi:10.1016/j.socscimed.2009.08.032.

“Discrimination in Metropolitan Housing Markets: National Results From Phase I of HDS2000.” 2000. “Discrimination in Metropolitan Housing Markets: National Results From Phase I of HDS2000.” United States Department of Housing and Urban Development.

“World Urbanization Prospects, the 2011 Revision: Highlights.” 2012. “World Urbanization Prospects, the 2011 Revision: Highlights.” New York: United Nations, Department of Economic and Social Affairs, Population Division.