ACCESS TO URBAN FOOD OUTLETS AS A PREDICTOR OF DIABETES

By

PHILIPPE AMSTISLAVSKI

A dissertation submitted to Graduate Faculty in Earth and Environmental Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
City University of New York

2012

This manuscript has been read and accepted for the Graduate Faculty in Earth and Environmental Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Juliana A. Maantay

Signature

Date

Printed name of Chair of Examining Committee

Dr. Yehuda Klein

Signature

Date

Printed name of Executive Officer

Dr. Luisa N. Borrell

Dr. William Burch

Dr. Yehuda Klein

Dr. Dana Charles Tomlin

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

ACCESS TO URBAN FOOD OUTLETS AS A PREDICTOR OF DIABETES

by Philippe Amstislavski

Advisor: Professor Juliana A. Maantay

Background and problem statement: There is an unprecedented rise in diabetes in urban populations worldwide. A relationship between spatial concentration of other metabolic diseases and poor access to healthy foods in some underserved urban neighborhoods have been reported. Concurrently, a relationship between increased risk of developing diabetes and consumption of unhealthy foods and has been shown to exist. Neighborhood food contexts hypothesized to lead to developing diabetes need to be studied.

Study goals: The main hypothesis of this study is that the degree of access to food outlets near residences influences the outcome of diabetes. Covariates include individual-level variables of age and gender of the subjects, and neighborhood-level variables of educational attainment, percent of residents in poverty, of housing units without vehicles, and of female-headed households with children.

Methods: Address, demographic, and health data extracted from medical records of black visitors to hospital emergency department were linked to geo-referenced socio-economic and food outlet data for the visitors' Census Tract (CT) of residence. Hierarchical linear modeling was used to consider the effect of variation in food and socio-economic environments on diabetes among the subjects. A cross-sectional study was designed and a multilevel logistic regression analysis was performed.

Results: Spatial access to food outlets was not a significant predictor of diabetes in this study. However, subjects living in the socio-economically deprived neighborhoods had a higher probability of having diabetes. For every unit decrease in the neighborhood's socio-economic index constructed from the census variables, the subjects were 7 percent more likely to have diabetes (CI 1.03-1.12, p-value 0.0024). Female gender and older age were strongly associated with odds of having diabetes.

Conclusions: Socio-economic context of neighborhood was shown to affect probability of having diabetes, while local food outlet access did not. The results indicate that there may be a critical difference between economic and spatial access to foods and the actual choices individuals make about their diets. These choices may be driven by individual cultural and social preferences. More research is needed to study these individual biosocial factors and to analyze how they affect diet and diabetes outcome.

Keywords: *Socio-Economic Environment, Diabetes, Energy Balance, Food Access, Multi-level Analysis.*

Acknowledgements

I am deeply grateful to my dissertation advisor, Dr. Juliana Maantay, whose devotion to her students and commitment to environmental justice provided an invaluable role model. Her kind encouragement, guidance and unwavering support enabled me to develop an understanding of the subject of this work and to carry out the research.

To Dr. William Burch, who trained me in social ecology at Yale, taught me to think out-of-the-box and to leverage my nursing experience and architect training to bear on the real-world problems of environmental justice and public health. It is thanks to his urging that I dared to begin exploring human health in its full complexity.

To Dr. Dana Tomlin, who facilitated and encouraged my interest in spatial analysis whose pioneering insight, guidance and sense of humor have enabled me to carry out several critical parts of this study.

To Dr. Klein, whose steady hand guided me through the labyrinth of crafting an inter-disciplinary doctorate at the Graduate Center of CUNY.

To Dr. Luisa Borrell, whose kind support and valuable advice on the intricacies of SAS code were crucial at the analysis stage of this study.

I am also thankful to Drs. Richard Sinert and Arun Subramanian at the Emergency Department of Kings County Medical Center. Without your help I could not have gotten such relevant clinical data.

I am in debt to all these who assisted me in developing data management and analysis schema for this study. Your names are too numerous to list here but I would especially like to thank Masayoshi Oka and Prakash Nadkarni, and my colleagues and mentors Paul Landsbergis and Jeanne Stellman at SUNY Downstate. Without your intellectual generosity and your dedication to improving the health of the public this work cloud not be completed.

Many dear friends shared the years during which ideas expressed in this study took shape. They included Cesar Moran, Boris Kinkulkin, Angelika Zollfrank and Sascha Manteufel. Your friendship and your ability to feign a continuing interest in my topic and your sense of humor have been invaluable.

Finally, I wish to express my infinite gratitude to my wife Tashia, my parents Tatiana and Alexander Amstislavski and to my sister Olga for their unconditional love, dedication, and their intense sense of beauty, truth, and justice. To them I dedicate this work.

Table of Contents

Abstract iii

Acknowledgements v

List of Figures x

List of Tables xi

Chapter 1: Introduction and Background 1

Chapter 2: Framework for the Study of Diabetes in an Urban Environment 11

- Introduction 11
- Effects of Neighborhood SES on Energy Balance 13
- Environmental Factors and Diabetes 20
- Change in American Diets and Diabetes 22
- The Effect of Diet Changes on Underserved Populations 23
- Modeling Approaches 25
- Other Methodological Challenges: 27
- The Role of GIS in Neighborhood Health Research 31
- Framework for the Study of Diabetes in the Neighborhood Context 32
- Chapter Summary 37

Chapter 3: Study Goals and Methods 38

- Introduction: Data Sources Overview 38
- Health Data Needs 39
- The EMR-Derived Database 43
- Geocoding of Subjects to Residences 45
- Selection of Cases and Controls 48

Definition of the Neighborhood Boundaries.. 51
Neighborhood Environment Data 53
Food outlets data 54
Neighborhood SES Data 65
Methodology for Computing Neighborhood SES and Educational Attainment Markers 66
Measure of Poverty 68
Measure of Private Vehicle Ownership 69
Percentage of Female-Headed Households with Children under 18 Years old 69
Percentage of Black Residents 70
Migration 70
Measure of Foreign Born Residents 70
Linguistic Isolation Markers 71
Chapter Summary 71
Chapter 4: Results 72
Introduction and Statistical Analysis Methods 72
SES Factor Analysis 74
Factor Analysis of Food Outlet Density 78
The Relationship between Diabetes and the Neighborhood Food Context 79
Aim [A] Measurement of Small Area Food Environments Around the Study Subject's Residence 81
The Creation of the Walkable Street Grids 82
The Calculation of Pedestrian Walking Speed and Distances 84
Characterization of the Food and SES Environment 90
Measures of Food Outlet Density 93
Aims [B]and [C]: To Characterize the Spatial Distribution of Diabetes Clusters in Brooklyn 99
Kulldorff's Spatial Scan Statistic 99
Aim [D]: To Measure the Effect of the Food and Socio-Economic Environments on a Diabetes Outcome 104

Multilevel Analysis 104
Chapter 5: Discussion, Limitations and Lessons Learned 111
Introduction 111
A Note on the Analysis Methodology 113
Diabetogenic Effects and Neighborhood Socio-economic Deprivation 114
Gender and Diabetes 117
Economic versus Spatial Barriers to Healthy Food 118
Limitations 120
The Challenge of Detecting Undiagnosed Diabetes Cases 121
EMR Extracts versus Population-based Data Sources 124
Health data continuity 125
Limitations in the Assessment of the Neighborhood Environment 128
Built Environment Effects 128
Modeling approach to Estimate Access to Food Environments 129
Future Research Directions 134
Importance of the Ethnomedical Model 135
In Conclusion 138
References 139

List of Figures

Figure 1.Map of diabetes prevalence in New York City 4
Figure 2. Proposed relationship between neighborhood SES, food environment and diabetes 18
Figure 3. Person and Place interaction in energy balance (After Daniel, Moore and Kestens, 2008).. 33
Figure 4. Spatial distribution of study objects 75
Figure 5. Spatial distribution of black population. 76
Figure 6. Residents without a High School Diploma 77
Figure 7. Percent of households in poverty. 77
Figure 8. Composite SES deprivation score map. 78
Figure 9. Raster image at 5 and 10 feet resolution 85
Figure 10. CWD calculation for supermarket access.. 89
Figure 11. The maps show kernel density estimates of food outlets using a .5 mile bandwidth.. 93
Figure 12. Graph of food density all types of licensed outlets per CT 95
Figure 13. High Food Density Tract.. 97
Figure 14. Very low food density tract with roughly the same population. 98
Figure 15. Results of the five levels of statistically significant diabetes clusters aggregated to CT of residence 103
Figure 16. Example of network-analysis based foodshed 129
Figure 17. A street vendor in central Brooklyn. 133

List of Tables

Table 1. Data elements extracted from the EMR. 47
Table 2. Sources of food outlets data. 56
Table 3. Sources of food store classification 60
Table 4. Food service establishments 62
Table 5. Factor analysis results. 74
Table 6.Results of the factor analysis on food outlets. 79
Table 7. Socio-economic characteristics of CT of residence by diabetes status. 81
Table 8. Characteristics of the local food environment of subjects by diabetes status. 91
Table 9. Demographic and Housing Characteristics of Subjects by diabetes status 92
Table 10. Model 1 Effect of individual-level variables on the outcome of diabetes 105
Table 11. Variance Inflation Factors 106
Table 12. Model 2; Effect of individual-level variables on the outcome of diabetes with added neighborhood socio-economic deprivation measures. 107
Table 13. Model 3; Effect of individual-level variables and food outlet density variables on the outcome of diabetes. 108
Table 14. Model 4; Individual level and neighborhood level SES variables 109
Table 15: 10 most common primary diagnoses in ED visitors during the study. 123
Table 16.Example of network-analysis based foodshed. 129

Chapter 1: Introduction and Background

The rise of diabetes and other metabolic disorders in the United States has been noted as a crisis-level problem for several years (Goldman, 2004). Scholars and policy makers alike believe that there is an urgent need to bridge the current gap in knowledge between interactions with our environment and our health. One of the main obstacles researchers face is the lack of data on environmental factors suspected of having a direct or indirect influence on metabolism and endocrine health.

In 2004, Hilary Rodham Clinton acknowledged the diabetes crisis and shortage of information on its origins: "Endocrine and metabolic disorders such as diabetes and neurological disorders such as Parkinson's are on the rise. The disease clusters and increases in their rates have led many people to look at the role of the environment in determining health status. What we know now is that too often there is little information on exposure with which to understand causal effects between the complexity of our environment and the reported increase of various diseases" (Goldman et al. 2004).

Recent notable examples of spatial and temporal disease clusters with a suspected environmental and socio-economic linkage include clusters of respiratory disease and air pollution (Nuvolone et al. 2011; Gehring et al. 2002; Maantay 2007; Maantay, Tu and Maroko 2009) and World Trade Center exposures in 9/11 rescue and recovery workers and subsequent respiratory and mental health problems (Friedman et al. 2011; Udasin et al. 2011). These and other instances of spatial and temporal clusters of illness point to the strong role of environmental factors in the etymology of health problems. As a result of these findings many researchers are now focused on developing better methods to collect and use environmental and socio-economic data in health research, in the hope of illuminating how the total environment affects the health of populations. In light of the unprecedented rise in diabetes in urban populations worldwide (Alberti et al. 2005), such work is critical.

Clinically, diabetes is a manifestation of insulin resistance, a condition in which means the body cannot use insulin effectively. The main risk factors to developing diabetes include visceral obesity, dyslipidaemia (elevated low-density lipoprotein cholesterol and triglycerides), hyperglycaemia (elevated blood glucose), and hypertension. Other risk factors include age, gender, race/ethnicity and socio-economic status (SES). All of these well-known risk factors have been causally linked to energy misbalance and are commonly grouped together as Metabolic Syndrome. In addition to greatly increasing the risk of diabetes, Metabolic Syndrome also predisposes individuals to other deadly conditions including cardiovascular disease (CVD), stroke, and certain types of cancer (Alberti et al. 2005). In addition to shortened life expectancy, concomitant heart and renal disease, reduced quality of life, loss of wages, and disability, diabetes patients as well as their families suffer from the disease's pervasive and wide-ranging social impact.

In 2007, the prevalence of diagnosed diabetes across all of the city's five boroughs was 9.1% for New Yorkers in all body mass index (BMI) categories and for the first time had escalated above the national prevalence of 8.3%. While the national prevalence had remained largely unchanged over a number of years, in New York City it was on the rise (Van Wye et al. 2008). More than half a million adult New Yorkers have been diagnosed with diabetes (NYC DOHMH, 2007). According to the New York City Community Health Survey, excluded from this already extremely high figure is an estimated 200,000 New Yorkers who have diabetes but do not yet know it because it has not been diagnosed. In their study Van Wye et al. (2008) used two robust data sources, the New York City Community Health Survey (NYC CHS) and the Behavioral Risk Factor Surveillance System (BRFSS). NYC CHS has been conducted annually since 2004 and collects data from randomly selected groups of New York City adult residents through a detailed health interview and brief physical exam. It is administered as a random-digit-dial telephone survey in which approximately 10,000 New York City adults aged 18 years or older participate. The authors compared the estimates of diabetes in New York City collected by NYC CHS to the national measures, reported by the BRFSS. Diabetes and diabetes-associated cardiovascular disease recently became the two leading causes of death in New York City (Van Wye et al., 2008), and diabetes-related mortality rapidly increased by 71 % between 1990 and 2003 (NYC DOHMH, 2006) and continues to climb.

Spatial disparities seen in the national epidemic of metabolic syndrome-related illnesses, including diabetes, are also extremely evident in New York City. In some neighborhoods, the percentage of diabetes cases remained unchanged for over a decade, while in other neighborhoods, diabetes rates continued to rise rapidly (Thorpe et al., 2009a; 2009b). Pointedly, residents of medically and economically underserved areas with poor spatial access to healthy food choices, including northern Staten Island, South Bronx, Harlem, northern and central Brooklyn and parts of Queens, have the highest prevalence of diabetes (Thorpe et al 2009a; 2009b). The cross-sectional NYC HANES study of residents of all five boroughs found that the reported prevalence of diabetes varied by zip code and ranged from 4.1 % to 16 % with the poorest areas carrying the highest burden of the disease. The lowest income areas of East New York, northern and central Brooklyn, have some of the highest rates in the city (NYC DOHMH, 2007)[1]. The map in Figure 1 depicts these stark geographic differences in the prevalence of diabetes across New York City estimated by the NYC Health and Nutrition Examination Survey (NYC HANES) in 2008.

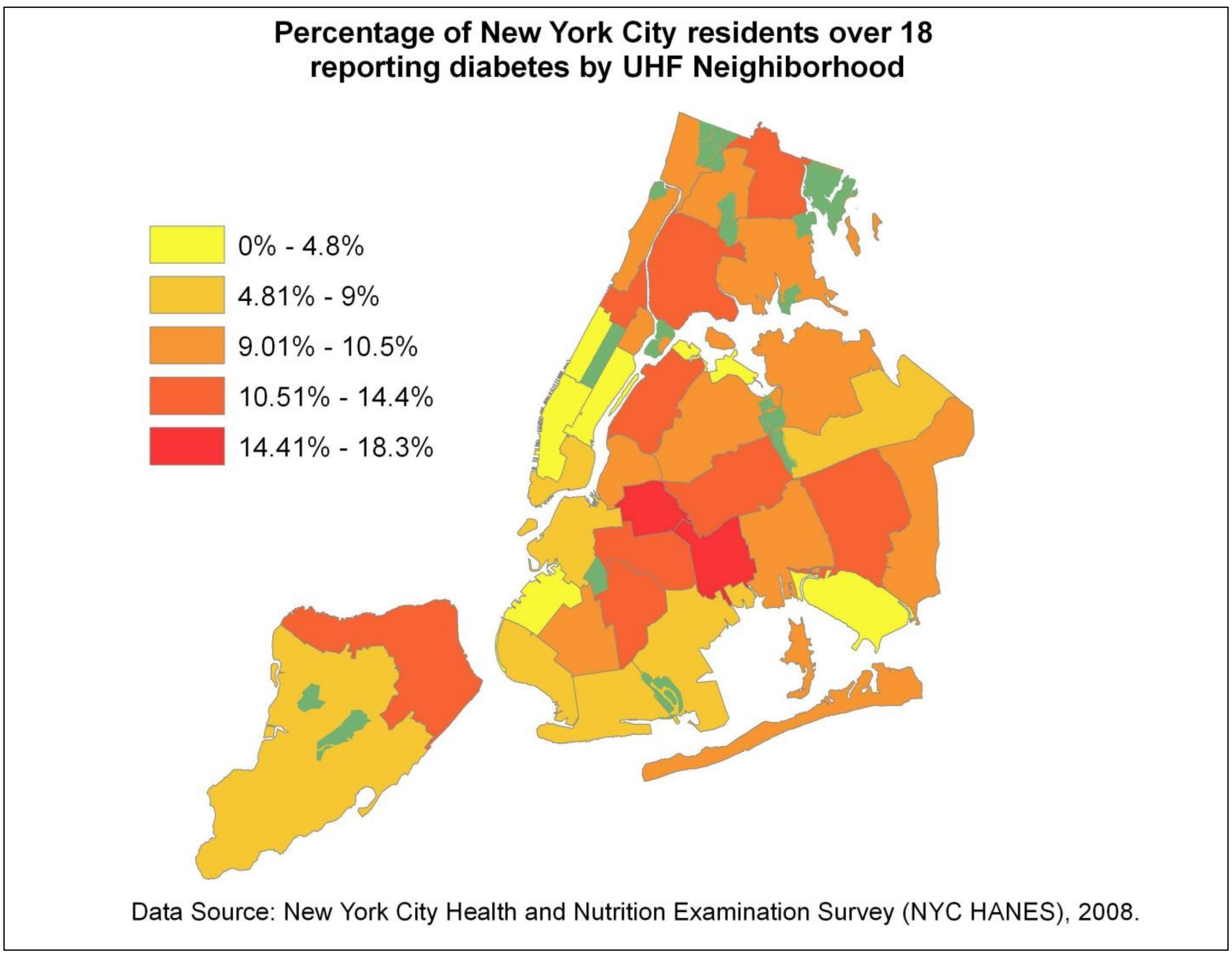

Figure 1. Minority and low income areas of New York City have the highest rates of diabetes.

The map of diabetes prevalence in Figure 1 was constructed using the data from the 2008 NYC HANES study, which divided the city into 34 United Hospital Fund (UHF) neighborhoods (NYC HANES, 2008). The 2008 NYC HANES study also reported wide disparities in the prevalence of diabetes among the city neighborhoods and noted that over 18% of adult residents of East New York and central Brooklyn, where the current study is being conducted, have diabetes[2]. Other studies conducted in New York City found similar results (Rundle and Neugut 2009; Thorpe et al. 2009a; Gwynn et al. 2011; Frieden et al. 2008; Althoff et al. 2009).

According to the NYC Community Health Survey, although the prevalence of diabetes in Brooklyn was 9.4%, lower than all other outer boroughs except Staten Island, its prevalence in the poor and minority neighborhoods of northern Brooklyn ranged from 13 to 16.9 percent (Frieden et al. 2008; Thorpe et al. 2009a). The higher rates of diabetes in parts of New York City prompted the former New York City Health Commissioner, Commissioner Frieden, to call this situation a diabetes epidemic (Frieden et al. 2008).

The findings of these studies point to an unprecedented increase in the prevalence of diabetes in New York City, and to growing disparities in diabetes. This situation has sounded an alarm in the public health community and has provided an additional motivation for this work, which centers on examining and explaining the linkages

[2] Based on the United Hospital Fund definitions of neighborhoods, which group several zip codes together based on their presumed relative socio-economic homogeneity. For a complete listing of all 34 neighborhoods and their zip codes, please refer to nyc.gov/health.

between local food access and neighborhood socio-economic context and diabetes prevalence.

Recent research has shown that in New York and other large and socio-economically diverse US cities the diabetes epidemic is most pronounced in areas with high concentration of various unhealthy food outlets and decreased opportunity for daily physical activity, such as parks (Khan et al., 2009). In other words, environmental factors, rather than biological or genetic ones, may be the key elements driving the diabetes epidemic in New York City. An increasing amount of literature on the relationship between obesity, diabetes and the built environments suggests that these illnesses may be related to the specific socio-economic and environmental characteristics of neighborhoods where the affected individuals live and work (Van Wye et al. 2008; Thorpe et al. 2009a; Lee et al. 2008). However, unlike obesity, diabetes has a long latency period, during which the impact of prolonged exposures to individual-level as well as neighborhood food context and socio-economic factors may contribute to other exposures that have been directly or indirectly linked to diabetes (i.e., low physical activity levels due to the lack of activity opportunities in the area, elevated level of psychological stress due to crime). In the case obesity, on the other hand, the body typically responds fairly quickly to the unhealthy environment and unbalanced nutrition by accumulating excessive amounts of weight. This key distinction in etiology of the two conditions accounts for the differences in association between the risk of acquiring the disease and the longitudinal exposures of the local food and socio-economic conditions affecting the local diets.

In summary, it can be stated that after years of exposure to the same contextual factors that lead to obesity, a slow, gradual impairment of glycemic control develops, which is why the so-called “twin epidemics” of obesity and diabetes often plague the same neighborhoods. As the body loses its ability to compensate for these offenders, the process eventually leads to diabetes. But, unlike in case of obesity, the multiple layers of exposures to social, behavioral, dietary, and environmental factors that have been for years hypothesized to increase the risk of diabetes often take place decades before the individual is diagnosed. In New York City alone, it has been estimated that over 3.8 % of all adults have undiagnosed diabetes and large percentage of these persons have been estimated to initially have developed obesity (Van Wye et al. 2008; Thorpe et al. 2009b).

Previous research has noted that environmental characteristics that increase the risk of metabolic disorders include the lack of reasonably priced grocery stores and restaurants offering healthy food. These degraded food environments are often positively associated with high level of poverty and with socio-economic deprivation (Gwynn 2011; Congdon 2010). The same conditions may also directly or indirectly influence the risk of developing diabetes due to the reliance of the local diets on the unhealthy local food environments. A study on New York City communities with a range of low to high on socio-economic deprivation characteristics, found that the prevalence of metabolic syndrome and diabetes ranged widely from 4.3% to 16.9% (Thorpe et al. 2009a) among the areas surveyed and was most associated with the degree of socio-economic deprivation in neighborhoods. Of special note, the neighborhoods with the highest prevalence of diabetes were also plagued with degraded food, transportation, and physical activity barriers (Ershow et al. 2007; Lopez-Zetina et al. 2006; Perdue et al. 2003). There is a growing consensus in the public health community that the local environmental and contextual features of where people live and work powerfully influences the choices that they make concerning their diets (Galvez et al. 2008; Morland et al. 2002; Morland and Evenson 2009; Auchincloss et al. 2008; Diez Roux 2009; Rundle et al. 2007). These conditions impact health behaviors and, through direct and indirect pathways, affect metabolism and influence the risk of developing diabetes.

However, the majority of the previous research on neighborhood context and metabolic disorders focused almost exclusively on obesity, often a precursor to diabetes, rather than directly on diabetes (Bader et al. 2010; Congdon 2010; Lee et al. 2008; Maroko et al. 2009; Michimi and Wimberly 2010). Previous research has shown that the prevalence of obesity and of the concomitant diabetes have been rising unevenly across the city, with the largest increase in prevalence in poor and underserved areas of New York City (Rundle et al. 2009).

The recent epidemiologic and medical geography literature highlights the effect of built, social, and economic determinants, such as housing conditions; food access and food security, income, and health care access inequality on individual behaviors and outcomes. Based on this framework, metabolic disorders, including diabetes, result from a confluence of these factors acting together upon an individual. Thus, variations in exposure[3] to these contextual variables, such as access to food outlets, may play a key role in the geographies of diabetes in urban areas.

This study examines the relationship between the geographic variation of diabetes in Brooklyn and food as well as socio-economic environments in Brooklyn neighborhoods. The impetus for this work was the high prevalence of diabetes in the central and northern parts of Brooklyn (Frieden et al. 2008) and the relative dearth of work on the contextual influences of diabetes and socioeconomic deprivation. Social and economic deprivation may create spatial conditions favorable to diabetes and produces spatial clusters of diabetes within the urban fabric wherever these conditions exist. In other words, spatial access to unhealthy food alone may not play a key role in driving the increased risk of diabetes. However, the neighborhood economic deprivation may create conditions where both spatial and economic access to healthy nutrition is decreased. Thus, economic status of the neighborhood and not physical access to particular types of food outlets may be of paramount importance in influencing the risk. This study attempted to provide an answer to the research question of whether poor spatial access to food outlets is statistically associated with increased diabetes prevalence if other relevant factors, such as social and economic conditions in the neighborhoods surrounding the subjects' residences are held constant. In Chapter 2 I focus on the theoretical framework of this research and its data sources; Chapters 3 and 4 describe analytic methods and findings, and their meaning in

[3] This characterization of the environmental exposure is distinct from the traditional notion of exposure in environmental health and environmental epidemiology, where it typically refers to exposure to toxic compounds.

the context of the improvement of urban health. The dissertation concludes with the discussion on lessons learned and future research directions outlined in Chapter 5.

Chapter 2: Framework for the Study of Diabetes in an Urban Environment

Introduction

This chapter explores some of the social underpinnings of the concept of health and place as it relates to energy balance, the state in which the total energy intake equals total energy needs, and particularly focuses on diabetes. I discuss an emerging framework for studying neighborhood health. This framework, which emerged from the nascent field of urban ecology, was extremely instructive in the current study. I also discuss diverse methods used by other investigators to measure the effects of socio-economic environment and food access on diabetes and other metabolism-related disorders.

The ecological framework for understanding place-based health and health behaviors suggests that health status is an amalgam of multiple levels of influence from the individual level to the neighborhood level and beyond (Auchincloss et al. 2008; Daniel, Moore, and Kestens 2008; Nuvolone et al. 2011). These levels expand concentrically from individual genotype, to biological, behavioral, social and cultural influences, and from family and friends to the neighborhood, regional and even societal level influences. For example, the suspected gene mutation responsible for insulin resistance in diabetes mellitus has been shown to result from poor diet (Trucco 2009), and it can be argued that such individual level effect is caused by the low social and economic standing which perpetuates poor access to healthy foods. Both epidemiological and basic science studies have also shown similar mutagenic effects of the bio-social environment in diabetes as well as several types of cancer (Chen et al. 2002; Groop & Orho-Melander 2001).

This chapter constitutes a review of the seminal peer-reviewed literature on the effects of different facets of the urban environment on the body's metabolic balance in general and focuses on the effects of these factors on diabetes, specifically. The discussion begins with a review of the effects of barriers to healthy food access on diets, and then proceeds to include the neighborhood's socio-economic status. The meaning socio-economic status as it pertains to this paper is 'the social and economic factors that influence what position(s) individuals and groups hold within the structure of society (Lynch and Kaplan 2000). Within the context of the relationship between diabetes and neighborhood, I discuss the utility of a socio-medical model of metabolic disease in respect to the study of diabetes.

In addition to the impact of social and food environments, exposures to certain pollutants, thought to act as endocrine disrupters, have also been hypothesized to be related to diabetes. While the nascent research on suspected endocrine disrupters and diabetes is in its infancy, I discuss the recent work in this area.

The ecological framework that was adopted for this study posits that the individuals are nested within the larger context of their neighborhood environment. After further review of the literature and data collection for the current study, it became apparent that to analyze the rich health outcomes data being collected at the local hospital and to test the effect of the spatial food access and neighborhood SES of diabetes, one would need to employ analytic techniques that are responsive to the hierarchical structure of the data. This is because such analysis needs to lend itself to testing the effects at multiple levels of influence. Therefore, this chapter also reviews hierarchical or multi-level modeling methodology, an approach often used for statistical modeling of spatially referenced data where independent variables are nested hieratically, within several levels of influence. In the case of this study, individual or subject-level variables and variables acting on the larger scale, such as those affecting the entire neighborhood of the subject's residence, are included in the analysis. The results of this review then led me to develop a cross-sectional study and perform a multi-level analysis of the data, discussed in depth in the following chapters.

Effects of Neighborhood SES on Energy Balance

In the last decade, the suspected relationship between low SES, unhealthy food access and food insecurity, and various metabolic spectrum diseases has come to the forefront of health research and policy (Daniel et al. 2008). The nature of the relationship between SES and the degree of access to healthy food and its effect on diets has also been highlighted (Gundersen et al. 2008; Dinour, Bergen, and Yeh 2007; Crawford et al. 2007; Martin and Ferris 2007; Morland et al. 2002). Work on this relationship is cross-disciplinary and encompasses the fields of epidemiology, medicine, psychology, urban planning, geography, and sociology, among others. The major limitation of this earlier research, however, was its main focus on the relationship between food access and obesity in urban areas. As a result, the effects of neighborhood SES and food access on diabetes, a much more severe condition, have been left relatively understudied, which was important to the focus of this study.

The previous work has shed some light on the complex relationship between neighborhood deprivation and diabetes and has shown that race and ethnicity are also very influential in the risk of developing the disease. (Gwynn et al. 2011; Bader et al. 2010 ; Ettner et al. 2009; Galvez et al. 2008). This effect is not necessarily because of the genetic predisposition to diabetes, but probably due to race/ethnicity acting as a proxy for geographic location. This influence, coupled with the effects of segregation imply neighborhood effects, rather than race/ethnicity-based causes of diabetes risk. While recent studies have also demonstrated that there is a relationship between the degree of spatial access to various food outlet types within the built environment and a range of health outcomes related to metabolism, the results are often mixed and conflicting (see for example Holben and Myles 2004; Whitaker and Satin 2007). Inconsistent results about which variables have protective or detrimental effects on diabetes made it a challenge to select independent variables for the current study, a theme discussed to some length in the following chapters.

In the late nineties, when the nascent quantitative research on neighborhood health first came to the forefront of the public health agenda, Schwab and Syme argued that in order to be successful in investigating health on the neighborhood level, one must “reflect the ecological reality of life in that population, as people experience it” (1997: page 632). Furthermore, Morland, Diez-Ruex, and others have called for further health research that acknowledges the importance of the differences both within and between neighborhoods

(Diez Roux 2008, 2009; Galvez et al. 2008; Morland & Filomena 2007; Morland & Evenson 2009).

One of the overarching conclusions based on the literature review is that social deprivation negatively influences energy balance through a number of pathways, thus possibly increasing the risk of both obesity and diabetes. When using multi-level analysis for neighborhood SES, reasonably consistent, but modest associations exist between residential area SES disadvantage and metabolic disease mortality rates (relative risks <2.0), and less reliably, morbidity rates (rate ratios close to 1.5) (Diez Roux 2009; Morland et al. 2002; Rundle et al. 2009). These research findings substantiate the claim that there is an association between neighborhood SES and ill health. Unfortunately, studies that show the strength of the association between ill health and place are modest and typically based on single-variable SES indicators, such as, aggregate income or educational attainment and utilized single-level, ordinary least squares (OLS) regression analyses (Daniel, Moore, and Kestens 2008). Expectedly, there are some concerns about the validity of such overly simplistic approaches used in the analysis of such complex health phenomena. The use of a single SES indicator as a proxy for such a complex set of conditions which create a particular social and economic climate in a neighborhood is often an inadequate measure of the total neighborhood SES and its degree of SES deprivation.

I suggest that a more analytically robust approach would be to use several SES indicators in order to construct an index of the neighborhood's socio-economic deprivation. Socio-economic deprivation indices are composite SES measures, which provide an approach to conceptualizing and measuring the broader construct of the socio-economic status of a neighborhood. They are widely used in health research in some European countries and often combine occupation-based and education-based socio-economic measures (Kirkpatrick and Tarasuk 2010; Longley and Singleton 2009; Sharkey et al. 2009). Although socio-economic deprivation indices have been used extensively by researchers in Europe and Asia for several decades, such indices have not been well developed for the US and therefore are not widely used (Gangwisch et al. 2007).

In order to measure the effect socio-economic deprivation, a simple socio-economic deprivation index was developed and tested for Brooklyn in order to make use of the more sophisticated Hierarchical Linear Model (HLM).

HLM is a type of regression model increasingly used in neighborhood health research (Auchincloss et al. 2008; Chen et al. 2005; Ettner et al. 2009; Rundle et al. 2008). HLMs have been used extensively in the analysis of the multifaceted effects of the environment on health. HLMs are seen as an improvement to such traditional techniques as ordinary least squares and other "single-level" regression methods because one can examine the effects of a number of independent variables that are nested within different levels of the model. Unlike, OLS regression analysis, HLM permits us to derive effects (or regression coefficients) per neighborhood in the study area. If we use the example of poverty effect on health, we can obtain the poverty effect for all the neighborhoods in the study where subjects live. The key advantage of HLM is that the researcher can weigh the neighborhood poverty effect on the outcome of interest and report the relative precision (or reliability) of this effect. HLM permits the reliable measurement of the neighborhood-level poverty effect in the calculation of the regression coefficients.

Despite the urgent need to define neighborhood-level factors that have an influence on the continuing increase of diabetes in some areas, a thorough review of the literature revealed a scarcity of studies that used HLMs to study the effects of SES and food access on an outcome of diabetes in NYC and globally.

Although more recent studies led by Rundle (2009), Morland (2008), and Diez-Roux (2007, 2009) examined the effects of the neighborhood environment on obesity and not on diabetes, they still provided valuable methodological contributions to this research. Similar to obesity, the commonly proposed medicament to curbing diabetes by the public health officials is to incorporate healthy foods and daily physical activities into the lifestyles of inner city residents. However, it may be impractical for residents of these disadvantaged areas to do so because of numerous barriers to changing diets and lifestyles. In order to remedy the situation, we must first delimit both the offending factors in the local environment that are associated with diabetes and the factors that have a positive effect on energy balance and may be protective against developing the disease. Figure 2 proposes a set of relationships to be analyzed using HLM to better understand the social and food environment antecedents of diabetes in the disadvantaged areas, such as the study area in this research.

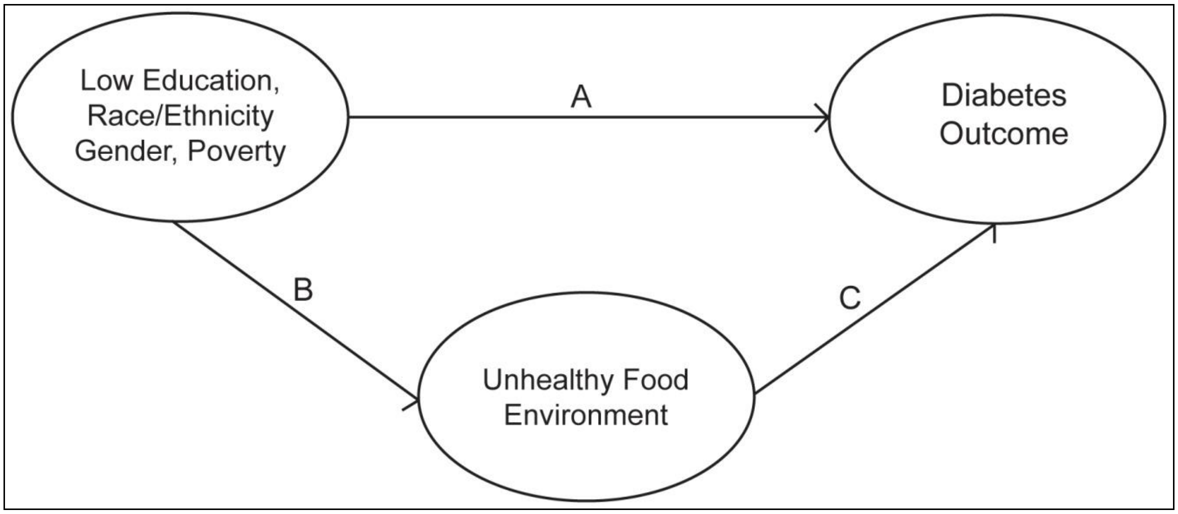

Figure 2. The proposed relationship between neighborhood SES, food environment and diabetes

With few exceptions (Rundle et al. 2007; Morland et al. 2006), most studies relating the food environment to diet or metabolism focused on the few selected types of food outlets rather than on the entire foodshed of individuals in a given community (Galvez et al. 2008, Hurt et al. 2010, Sharkey et al. 2009). Because the density of different vendors of the same type of food are likely to covary with each other, with the competing vendors of other types of food, and with commercial space availability in general, the spatial relation and density of these outlets may also have an effect on their contributions to the total foodscape and food access. But since most studies focus exclusively on one or a few types of outlets (e.g., fast food chains or supermarkets), the effect of these covariates will be very difficult to separate. This limitation in the prior research prompted me to study the effect of the total food environment by including all known food retail outlets in the analyses.

Prior work on the linkages between neighborhood environment and diabetes used a spatial analysis methodology seen as problematic in measuring access to food sources. Traditionally, various types of buffers are set up around the subject's residence, which are then used to measure access to food outlets. In these earlier works the food outlets captured in the measurement of the food environment, were typically defined as falling within these spatial buffers (Rundle et al. 2007; Rundle et al. 2008; Rundle et al. 2009; Auchincloss et al. 2008; Morland and Evenson 2009) or within an arbitrarily assigned "neighborhood", usually defined by Census tabulation units or other administrative districts (Morland and Evenson 2009; Galvez et al. 2008). However, neither approach sufficiently measures the walking access to various food outlets. Walking is perhaps the most common means of seeking food in the densely populated borough of Brooklyn (Tudor-Locke et al. 2009). The network analysis method can be seen as a "catch-all or none" approach because it only assesses whether points, representing food outlet locations, fall on the inside or the outside of the area defined as the subject's "neighborhood". But the sharp boundary buffers constructed in GIS by researchers, or by the Census tabulation units have no effect on food shopping in real-world conditions. In fact, in densely populated and socio-economically diverse urban areas, such as New York City, the individual's food environment may be radically different from one city block to another. But when we measure "exposures" to different kinds of food outlets by aggregating the food outlets into larger, administrative districts (such as Census Tract or CT) or network buffers (i.e., using network buffers with study subjects at the center) this block to block complexity and variation will tend to disappear. Due to this limitation in the current methods, this study sought an alternative methodology for measuring spatial access to food outlets, which is described in Chapters 3 and 4.

An additional issue that emerged during the review of earlier studies on various types of energy balance disorders and food environment, was the quality of data on businesses selling food. In most food environment studies referenced in his chapter, the food outlet data was derived directly from commercial US business listings companies, such as Dun and Bradstreet Corporation or InfoUSA, Inc. These large companies collect and sell business listings data, mainly to entities engaged in market research. However, some researchers suggest that such companies often do not collect listings for the smallest, independently-owned businesses, such as bodegas and corner stores (Lavin 1998). Yet such small, independently-owned businesses often constitute the most ubiquitous type of food outlet in some urban areas. Therefore the commercially available business listings from companies, such as Dun and Bradstreet and InfoUSA, may be insufficient for research because listings of the smaller, independent food outlets may be missing. Therefore, an alternative source of data to estimate the food environment was used. This important topic will be discussed in some detail later in this chapter.

Environmental Factors and Diabetes

There is a growing concern that in addition to diet and activity levels, toxins in the home or in the outdoor neighborhood environment may act as endocrine disrupters and thus play an important role in the development of diabetes (Hotchkiss et al. 2008: Van Den Hazel et al. 2006). For example, one regional study in the Hudson River watershed (New York State) linked the presence of persistent organic pollutants (POPs) and their effect on insulin resistance and diabetes among Upstate New York residents (Kouznetsova et al. 2007). Lee at al. (2007) also reported a strong association between high concentrations of several suspected endocrine disrupters—namely a group of POPs —in the blood and the high prevalence of diabetes among U.S. populations with a background of exposure to POPs, such as farm workers. The POPs studied by Lee's team are organochlorine (OC), pesticides, and non-dioxin like polychlorinated biphenyls (PCBs). Lee's findings, while examining diabetes in American farmers, indicated that exposure to the POPs can be associated with insulin resistance, which is a strong pathogenic precursor of type 2 diabetes.

The 2007 Kouznetsova at el. study examined the hypothesis that residential proximity to POP-contaminated waste sites in Upstate New York resulted in increased rates of hospitalization for diabetes. To assess the risk of acquiring diabetes due to the possible exposure to PCBs, Kouznetsova et al utilized the New York Statewide Planning and Research Cooperative System (SPARCS) data on diabetes diagnosis which was aggregated by zip code. Unfortunately, this otherwise important study suffered from a methodological flaw. Kouznetsova et al. used SPARCS data on hospitalizations only for the residents of zip codes near the hazardous waste sites in the Hudson River watershed, already contaminated with PCBs due to General Electric's activities. The study concluded that there was a strong association between the proximity to PCB sites and diabetes—individuals with high levels of PCBs in the blood had a 3.9-fold (95 % CI 1.5—10.6) chance of having diabetes. However, this research has been criticized due to the suspected ecological fallacy. The criticism stated that the link between PCB's and diabetes could not be proved because the study did not control for other factors linked to diabetes. Studies measuring the levels of PCBs in the blood, prevalence of diabetes, and that control for other confounding factors are needed to further investigate the linkages between PCBs contamination and diabetes. Future studies measuring biomarkers, such as levels of PCB in the blood, in diabetics and controls in New York State (with its high prevalence of diabetics in certain areas), are required to shed more light on these suspected linkages.

Change in American Diets and Diabetes

The National Health and Nutrition Examination Survey reported that across all population groups measured the prevalence of diabetes tripled between 1980 and 2007, and overweight and obesity (two conditions linked to the risk of developing diabetes) also increased from 47% in 1980 to 65% in 2002 (Gangwisch et al. 2007). It seems implausible to suspect that specific genotypes alone are responsible for such an unprecedented climb in diabetes rates within a very short time span. This research and other studies suggest that the rapid rise in the prevalence of diabetes throughout all populations in the U.S. cannot be explained away by implicating the role of genotype alone, but, the continuing increase in the number of calories Americans consume daily, should also be considered (Obesity Action Coalition, 2007; White, 2007).

The increased portion sizes across all major food groups in the American diet are concomitant with the serving size increases in food stores and food service establishments. Researchers have made a link between the massive advertisement campaigns of high-caloric and low nutrition foods and the increased reliance of Americans on foods served outside the home, and the prevalence of diabetes and obesity (Gamble and Cotugna, 1999). It was estimated that in 2007 40% to 50% of every food dollar was spent on food outside of the home (Deshpande et al. 2008; Ershow et al. 2007). It is reasonable to expect that food consumed when eating out is not as healthy as the food made at home.

Childhood obesity, specifically, is acknowledged as being one of the major pathogenic precursors of diabetes. A paper by the Obesity Action Coalition, a non-for-profit organization dedicated to curbing obesity in the US, reported that sugared beverages, such as soda and juice boxes, contribute to childhood obesity (DiNapoli and Lewis 2008). The consumption of soda by children has increased throughout the last 20 years by 300%. It is estimated that 20% of all overweight children are overweight due to excessive caloric intake from soft drinks and juices.

The Effect of Diet Changes on Underserved Populations

The detrimental effects of unhealthy food may be directly or indirectly related to the poor quality and low choice of healthy foods available in neighborhoods with degraded economic environments where individual’s live and work. There is strong emerging evidence that local SES affects spatial and economic access to healthy food and thus influences the daily dietary choices made by individuals living in disadvantaged areas. Specifically, studies showed that low neighborhood SES is strongly related to the low consumption of fruits and vegetables and the high consumption of fats and carbohydrates in poor neighborhoods (Stotts Krall and Lohse 2009; Park et al. 2008; Morland et al. 2002).

There is ample evidence in recent epidemiological and biomedical literature that a strong positive relationship exists between diabetes and unhealthy diets. (Bader et al. 2010; Hurt et al. 2010; Adamo and Tesson 2008; Galvez et al. 2008; Pedersen, Kang, and Kline 2007; Schröder 2007). These studies, coupled with the NYC HANES data led to the study's initial hypothesis that poor access to healthy food options in underserved urban communities may increase risk of acquiring diabetes by the residents. The conditions which produce degraded local access are only exacerbated by residents' limited mobility and low buying power. Thus the residents of the disadvantaged areas become "captive shoppers," their food choices limited to the few options available in the outlets around their homes. Children and seniors, known to have the least spatial mobility due to limited access to private transportation, may have the most restricted food choice options. Likewise, poor urban households, without private vehicles, are restricted to mass transit or walking, and have fewer shopping choices. Therefore, poor urban minorities may be greatly affected by these spatial constraints and be limited by their local food environment.

For example, the New York City neighborhoods that have the lowest density of supermarkets, stores that offer the most food variety, are typically those that are poor and predominately black. Morland et al. (2002) demonstrated that white populations have greater mobility compared to inner city black populations. As a consequence of limited mobility, the authors reported, these individuals may be forced to rely solely on the food environment that is near their home. In short, if healthy foods are absent from the local shelves, so is the opportunity to eat healthy.

Modeling Approaches

Most health and epidemiological data are spatially-referenced. Such place-linked data needs specialized analytical techniques. These spatial analysis techniques take into account the fact that the observations of spatially-linked events are not independent, a key assumption of the classical statistical methodology. Health data is typically derived from observations of events that are spatially autocorrelated, which would not be the case if the events were randomly located over the cartographic space. Several approaches have been developed to analyze such spatially-referenced health data, which are not covered by the traditional statistical methods.

Geographically-weighted regression (GWR) and HLMs were specifically designed to address this gap. Geographic and spatial statistics literature has responded to the challenge of developing analytical tools to better understand health and health behaviors in place by applying sophisticated methods of hierarchical or multi-level statistical analyses (Chen et al. 2005; Subramanian, Huijts, and Perkins 2009; Subramanian and Kennedy 2009; Tucker-Seeley et al. 2009). Below is a brief overview and critique of these methods in light of this study.

Hierarchical or multilevel structures are now the norm in the study of the environmental effects on health. In these statistical models, individuals can be viewed as nested within geographical areas of residence or within their workplaces. Conceptually, we may expect that individuals with a particular health outcome will be clustered in specific areas because we assume that the health status of two randomly selected individuals from the same area of residence will be more alike than the health of two individuals selected from different areas of residence. Such isomorphism, also termed spatial autocorrelation, can be explained by the fact that the former pair is more likely to share similar environmental exposures, while the latter may not.

To demonstrate this logic, consider the following example. Children spend a large part of their waking hours in and around their schools. Specific features of the school, such as its design (e.g., number of stairs, classroom location and shape, etc.), available food options, availably of outdoor and indoor play space, and socialization to physical activities, all influence energy balance and thus may affect obesity in children. Because of the heterogeneity of these environmental characteristics among the different schools, we expect the obesity levels among children in the same school to be more similar than obesity levels of children from different schools.

HLM has been developed in part to analyze such geographically-linked phenomena within a hierarchical structure. In the school example above, the lowest level of observation in the hierarchy (e.g., child) would be the 1^{st} model level, and group or cluster (i.e., each school studied) would be the 2^{nd} level. A growing group of researchers has been employing the HLM analyses to study the effects of environment on health. For example, Subramanian et al. (2004, 2009) used HLM to demonstrate that subtle yet persistent contextual effects exist on many health outcomes across a diverse mosaic of biosocial landscapes.

Other Methodological Challenges:

Correlation versus Causality in Neighborhood Effects on Health

There are several key challenges that researchers face when analyzing linkages between place effects on health. One of these problems is confusion between causation and correlation. Many studies examining the effect of place on health and employing geographic information systems or GIS, assume a causal relationship between the presence of a particular contextual factor in the environment and the development of a specific illness in the neighborhoods where such contextual factors are found. But the mere presence of the contextual factor in the neighborhood does not automatically imply a causal relationship, i.e. the presence of a contextual factor may not actually cause the disease. The disease may be caused by other environmental factors, or a combination thereof, genetic predisposition to develop the disease, or an occupational exposure. Without further supporting data proving the causality of the exposure—disease link, we can only infer that there is a correlation but no causation. However, simply assuming that the disease is caused by the contextual factor without measuring and controlling for other factors that may be linked to the disease would be erroneous. Researchers should also be cautious not to hastily accept the null hypothesis, inferring that no causal relationship exists (Helms, 2008). The presence of the correlation may point to a causal link; but absence of evidence does not equate to evidence of absence.

Ecological Fallacy and the Modifiable Areal Unit Problem

Ecological fallacy and the Modifiable Areal Unit Problem (MAUP) are two important statistical problems that are key to geographic and health research. The concept of ecological fallacy can be described as a false inference about individual-level outcomes that were computed from a geographically aggregated sample. Such aggregation (to neighborhood, district, or some other arbitrarily chosen spatial extent) cannot be directly attributed to any individual that lives in the study area. It is not uncommon, however, to encounter errors in analytic reasoning that are rooted in ecological fallacy, especially in cross-sectional studies.

Until recently, much of the evidence on the associations between various exposures and neighborhood was based on analyses using patient-level outcome measures, such as BMI differences among subjects living in different neighborhoods or the odds ratios of a specific diagnosis based on the neighborhood of residence. (Longley and Singleton, 2009). The neighborhood effect of social and other contextual variables on the health of individuals has come to the forefront of the public health agenda. Concurrently, with the advent of the computerization of electronic medical health records and GIS, it has become fairly easy and inexpensive to join residence and work addresses of subjects in studies to the aggregated SES data for census tracts or similar administrative units where these individuals reside (Daniel et al. 2008). Socio-economic and demographic data for such research cannot be extracted from the health records and is not generally collected directly from the study subjects but rather is appended to the individual health outcomes from national censuses and other databases providing aggregate SES data. The census data is typically aggregated to various areal units, such as Census Blocks in the U.S., Communes in France, or Enumeration Districts in the United Kingdom. This "proxy" SES approach is problematic for two reasons: (1) given the unclear correspondence (isomorphism) between administrative boundaries of a place and of the social construct of "neighborhood" (MacIntyre et al. 2002), the administrative boundaries of data aggregation may not reflect the true neighborhood effect; and (2) the presence of a particular set of contextual variables cannot be assumed to equate to the same level of exposure across individuals from different socio-economic, racial, and ethnic groups or personal behavioral characteristics residing in the same neighborhood. This last point is important because these characteristics may be mediating the effect of the exposure in question. For example, effects of indoor air pollution may be mitigated by social status because individuals living in poverty may be exposed to higher levels of indoor air pollution due to poor quality of their housing units.

MAUP is another potential source of error that can affect health research that is reliant upon aggregate data sources (Unwin, 1996). Geographical data is often aggregated to areal units that reflect some type of administrative or political nomenclature. Boundaries of the polygons of such areal units, however, are often purely arbitrary in their nature. It is this arbitrary variation in areal resolution that generated the term 'modifiable'. For example, it could be argued that UHF neighborhoods in New York City, which were originally designed to demarcate racially, ethnically, and socio-economically homogenous areas of New York for health research purposes, no longer represent such areas accurately, or in a manner adequate to measure variables such as income. Instead, aggregation by housing type may be a much more appropriate approach because individuals living in public housing typically have lower incomes than those living in expensive private housing stock in the same UHF neighborhood. While the private condos may be located across the street from the housing project, “averaged” income data aggregated to a UHF neighborhood will smooth and thus mask income inequality within the UHF neighborhood. Maroko et al. (2009) developed a novel methodology to disaggregate spatial data, such as in the previously mentioned example, using cadastral datasets. It is anticipated that their approach will yield more accurate results which reflect the various income levels in such situations.

Large amounts of aggregated data require a careful choice of aggregation level to display the spatial variation of the data in a comprehensible manner. Ratcliffe and McCullagh (1999) provide an excellent overview of MAUP and its possible solutions for analyses of crime and SES data.

Ecological fallacy and MAUP are only two of several challenges which researchers face during the study of place effects on health. Attempts to fit spatial data into conventional statistical models often bears invalid results because the one fundamental requirement of conventional statistical data is that the samples are chosen and distributed randomly. Yet spatial data almost always violates this requirement (O'Sullivan, 2003) because of the issue of spatial autocorrelation. Individuals who live near one another tend to share many more of the same exposures and stressors than those who live in different places. This difference between traditional and spatial statistics is also reflected in collection methods commonly employed to gather spatial data because artificial boundaries often must be imposed on the area being sampled due to practical and subject privacy protection considerations.

The Role of GIS in Neighborhood Health Research

Mapping and spatial analysis offer the opportunity for researchers to visualize health data, generate and test hypotheses, and draw conclusions about the impact of place on health. In order to visualize and account for multiple layers of environmental influence, medical geographers and other health researchers have increasingly employed Geographic Information Systems technology. "Geographic Information Systems or GIS is, at its heart, a simple extension of statistical analyses that join epidemiological, sociological, clinical, and economic data with references to space" (Ricketts 2003). GIS is a rapidly expanding field of knowledge that exploits the recent advances in computer technology to organize, display and analyze spatial data. As such, it promotes better decision-making in the field of public health. In recent years GIS has become increasingly central to public health research in order to store, display and analyze a wide range of spatially-referenced health data. GIS can effectively be used to organize, visualize, and analyze spatial data, leading to a deeper understanding of spatial patterns of exposures and outcomes at various scales.

GIS has been widely used in health research for the visual exploration and presentation of health data; for identification of disease clusters and hot spots; in syndromic surveillance; distance estimation to evaluate health care access; and spatial analyses that investigate relationships between spatial distributions of health outcomes and potential causal factors, such as environmental contamination. Jarup (2004) wrote a comprehensive review of existing methodologies and techniques for disease mapping. One of his main conclusions is that although exposure and disease mapping using GIS modeling may enhance our understanding of the effects of the local environment on health, there exist several important limitations in respect to the quality of data used by such studies. Accurate and detailed records of health and demographic data are critically important for such studies. Jarup wrote that some outcomes data, such as cancer mortality and incidence, are well reported and of good quality, while other health data, such as birth defects, are reported poorly and inconsistently. Correspondingly, the International Statistical Classification of Diseases and Related Health Problems, 9th revision (ICD-9) codes for certain diagnoses, including congenital anomalies, abstracted from hospital admissions data, may be unreliable and of uneven quality (Jarup 2004). Additionally, like with any health data, the quality of hospital admissions data on a specific health outcome, such as diabetes, may vary between geographic regions and over time. Although geographic patterns of disease may theoretically be used to infer possible associations between environmental factors and their health effects, such patterns may not point to a specific association but actually reflect differences in health data recording.

Framework for the Study of Diabetes in the Neighborhood Context

As early as 1983, Colin Loftin and Sally Ward noted, in their seminal paper on the spatial analysis of socio-economic factors, that a fundamental principle of human ecology is that social characteristics are systematically and predictably distributed within the city (Loftin and Ward 1983). One would also expect this to be true in respect to this study, which focuses on examining the effect of food outlets on patterns on diabetes. However, diabetes is highly associated with not only with diets but also with the socio-economic conditions within a neighborhood and also with the individual SES. Therefore an integrative, hierarchical framework is needed in order to comprehensively model this multi-level relationship. Based on the work of several researchers, a useful framework has recently emerged for the analysis of SES and biosocial pathways for which the spatial clustering of disadvantage might be viewed as causally related to metabolism, cardiovascular, and glycemic disease (Daniel, Moore, and Kestens 2008). They have called for the adoption of this broad framework, shown in Figure 3, Person and Place Interaction in Energy Balance, in studying the associations between the biosocial environments and the spatial clustering of diabetes, which remains largely unexamined.

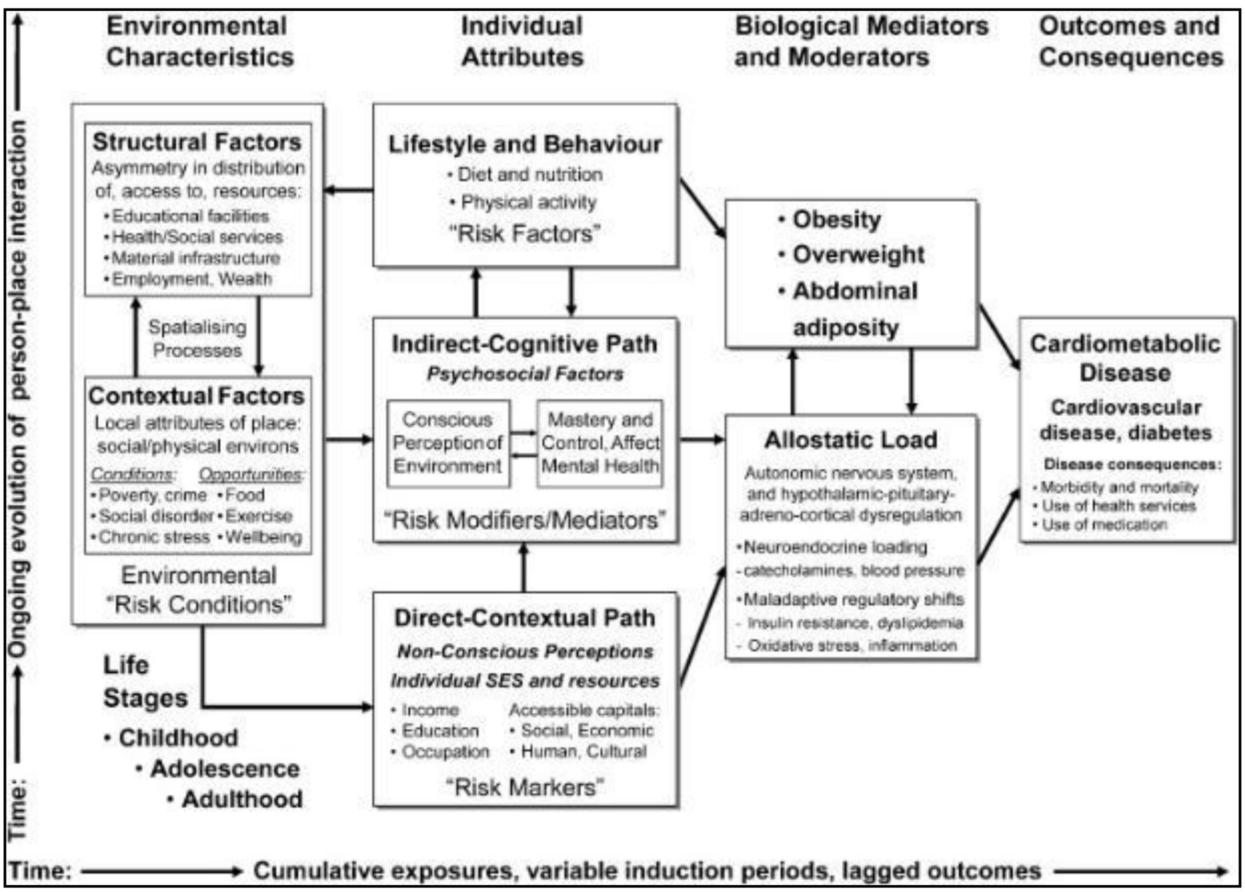

Figure 3. Person and Place interaction in energy balance (After Daniel, Moore and Kestens, 2008).

Daniel et al. proposed a diagram that is in many respects very similar to the Human Ecosystems Framework developed earlier by Burch and colleagues based on their work in Baltimore (Grove and Burch 1997; Field and Burch 1991). These complex amalgams of levels of influence included in the framework proposed by both Daniel et al and Burch range from social ecosystem level, and the neighborhood's constructed environment, to behavioral and cultural factors, and to genotype. Although these concentric levels of influence in the model lack a temporal scale, they do have a hierarchical order of influence. Importantly, these many influences can overlap with one another.

The theoretical biosocial framework proposed by Daniel is especially useful in this research. In the literature linking the neighborhood effects on health through socio-economic and biophysical pathways, it has been generally accepted that the availability of neighborhood resources that support healthy eating and physical activity are associated with the decreased odds of diabetes in residents (Larkin, 2003; Rao et al. 2007; Rundle et al. 2009).

The following hypothesis was tested by this study—whether spatial access to diabetes-unhealthy food outlets (these selling primarily processed foods with high-calorie, high-fat, low dietary fiber content and offering few food options) and diabetes-healthy outlets (these selling fruits and vegetables and offering a variety of food options) influences the probability of having diabetes when the aggregate neighborhood SES is held constant.

Based on the socio-medical literature on the effects of aggregate neighborhood socio-economic conditions of health, SES deprivation would also be expected to influence the local food environment and individuals' spatial and economic access to healthy foods and such relationship may not be linear. Thus, the food environment effect is seen as primarily driven by the neighborhood socio-economics. It is reasonable to hypothesize that the spatial segregation of persons with low education and low SES into neighborhoods that are poor in services and lack stores that offer healthy foods at reasonable cost creates conditions that may be causally linked to the early development of impaired glucose tolerance. In turn, glucose intolerance, a known precursor of diabetes, predisposes the residents to diabetes, often much later in life. In New York City and elsewhere in the U.S., spatial, class, racial and ethnic segregation creates conditions in which the quality of food available in these neighborhoods is consistently defined by the social class, race, and ethnicity of its residents. Not surprisingly, a disproportionately large percentage of metabolic diseases in general and of diabetes in particular, falls on poor New York City neighborhoods where most residents are members of racial/ethnic minority groups (Beebe-Dimmer et al. 2009).

Recently, the scarcity of healthy food choices in areas termed urban "food deserts" have been implicated in higher rates of metabolic syndrome in urban dwellers and have been hypothesized to act as a conduit to developing diabetes (Auchincloss et al. 2008; Lee et al. 2008; Wee et al., 2008). Morland and Everson showed, for example, that poor access to fresh fruits and vegetables contributes to increased obesity rates in poor New York City neighborhoods (Morland and Evenson 2009). Similarly, the very limited supply of and access to local healthy food and the limited physical activity opportunities that exist can be hypothesized to play an important role in the development of diabetes through several direct and indirect pathways. These pathways include poor access to healthy food and to exercise options, which may create new and reinforce already established behaviors that put individuals at higher risk of diabetes.

The differential access to healthy foods and the barriers to accessing it are clearly not limited to New York City. In fact, it is pervasive across American cities—economic, and behavioral barriers all play an important role in accessing food. For example, a recent study among low-income Pennsylvanians, assessing nutritional literacy, showed that cost, access convenience, mood, and the general availability of food in the neighborhood were limiting factors guiding meal and snack planning for the participants (Stotts, Krall, and Lohse 2009). Interestingly, nearly equal proportions of nutritionally literate and nutritionally illiterate persons (measured with the Satter diet literacy model) reported that they would make changes to their food purchases if they had more money to spend on food (Stotts, Krall, and Lohse 2009).

It is reasonable to think that the spatial disparities in food access put residents of disadvantaged areas at a disproportionately higher risk of developing diabetes because access to healthy foods, at reasonable cost, may be impeded for residents in poor areas, regardless of their individual SES, than those living in wealthier neighborhoods, who have easier access to healthy food choices. To date, sparse empirical data has been reported on the neighborhood effects of food and socio-economic environments on diabetes. This study aims to begin filling this important gap in knowledge.

Chapter Summary

This brief salient literature review underscores my overarching bio-social theory is that the complexity of the multifaceted interaction between individual's health and their urban environment must be appreciated and approached at multiple scales of analysis. Specifically, hierarchical data on various characteristics of the built and socio-economic environments of various neighborhoods that comprise Brooklyn, NY must be overlaid with individual level health outcomes data on a range of urban conditions that have been shown to be directly or indirectly related to metabolic imbalance. This is important because diabetes and of other metabolic syndrome-linked diseases are a result of an amalgam of social, economic, environmental, and genetic effects. The following chapter provides the reader with the methodology used in the acquisition, cleaning and coding of both contextual and individual health outcomes data in preparation for the analysis.

Chapter 3: Study Goals and Methods

Introduction: Data Sources Overview

This study seeks to quantify the effect of access to both healthy and unhealthy foods near residences on an outcome of diabetes, while controlling for the neighborhood-level variation in the socio-economic environments across Brooklyn, NY.

The key challenge to studying the role of environmental factors in diabetes and other metabolism disorders is the complexity of the interaction between individuals and the environments where they live and work. Such research often requires data from a large number of sources at different levels of spatial and temporal resolution on a range of environmental and social conditions that have been hypothesized or shown to be directly or indirectly related to the metabolic imbalance. For instance, to examine the possible role of the availability of different types of food in the urban environment on diabetes, accurate and spatially referenced data needed to be collected on both access to metabolism-healthy and unhealthy foods, and on the health outcome of diabetes in the study area.

To achieve the study goal, four principal aims were established: (1) characterize the total food environment in Brooklyn, NY, defined as all 9,739 licensed food stores and food service establishments operating in Brooklyn in 2008 (2) characterize the SES in the 452 CTs where study subjects live using US Census 2000 data; (3) use of a Brooklyn hospital electronic medical records database to develop a patient sample appropriate for a cross-sectional study (4) fit statistical models that predicts odds of diabetes based on access to different types of food vendors, as defined by aim 1.

Subject's residence address, demographics and diabetes outcome data was extracted from a hospital database. Since measures of socio-economic environments can be derived from the US Census data, aggregated to the Census Tract, this unit was used as a proxy for the subject's neighborhoods. Covariates in this analysis included the age and gender of the subjects, the percentage of Census tract residents with less than 9th grade education, the percentage with less than high school education, the percentage of residents living below the federal poverty line, the percentage of housing units without a vehicle, the percentage of minority residents, and of female-headed households with children.

Health Data Needs

There are several general problems shared by other researchers in acquiring high-quality environmental and health data, which were also encountered during this study. Firstly, the United States has no centralized disease registry that provides a collection mechanism for non-reportable diseases[4] such as diabetes (MMWR 2010). This is in contrast to most EU countries, which have national health insurance systems to facilitate timely health

[4] In the U.S. the Centers for Disease Control designate only cases of certain infectious diseases as reportable at the national level. A reportable disease is one for which regular, frequent, and timely information regarding individual cases is considered necessary for the prevention and control of the disease. Annually updated list of the nationally reportable diseases can be found on the CDC website http://www.cdc.gov/ Chronic disease cases, including diabetes, are not mandated to be reported on the national or state levels.

outcomes data collection into well-established national disease registries. While some states, including New York[5], do maintain health data collection systems, the technical issues with data quality and the administrative lag time required to access such health data repositories, often prevent researchers from timely access to data in state-run clinical data collection systems.

SPARCS versus Hospital EMR

This section addresses the complexities and relative merits of using two specific health data collection systems in the quantitative analyses, namely SPARCS and a hospital-based electronic medical records database. Outpatient data from the SPARCS database contains various data elements abstracted from the Emergency Department (ED) visit summaries submitted electronically by all hospitals in New York State. The diagnoses are recorded using the ICD-9, and thus in theory, the records of ED visitors with various diagnoses of interest can be extracted along with information on age, gender, race/ethnicity, name of ED, health conditions and their severity (Reilly et al. 2004). But patients rarely present to the ED with diabetes as the primary or secondary cause of the visit, thus this diagnosis may be underreported in the SPARCS. For example, the ED discharge summary provided to the SPARCS database of a patient admitted with trauma and a past medical history of diabetes would often only include ICD-9 codes with the trauma and related conditions, with no report of diabetes. Thus, this limitation in the

[5] See New York State SPARCS health data repository http://www.health.state.ny.us/statistics/sparcs/

SPARCS data presented a problem in accurately detecting all diabetes cases in the study population.

On the other hand, due to the recent advances in relational database management systems (RDBMS)[6], the EMR has become a widespread technology to record and archive the medical records of patients. In recent years, the EMR has therefore become a valuable source of data for research involving spatial analyses (Yiannakoulias 2009). While there is currently no federal standard establishing what data should be collected in the EMR, a typical hospital EMR contains a variety of health-related data, including ICD-9 diagnostic codes, medication information, laboratory tests, clinical notes, medical imaging results, demographical information, patient's addresses, and insurance information. With full access to a hospital medical record database, a complete search of a patient's medical history for all diabetes-specific ICD-9 codes, as well as other indications of diabetes diagnosis, such as mentions of diabetes history in the medical history notes, and laboratory results for elevated blood glucose values could be reviewed using SQL[7] queries.

[6] RDBMS is a type of database in which data is stored in the form of tables and the relationship among the data is also stored in the form of tables Codd, E. F. (1970) A Relational Model of Data for Large Shared Data Banks. *Information Retrieval*.

[7] SQL stands for Structured Query Language, a powerful and diverse database programming language. SQL can be used to create and query complex databases. Its simple syntax permits scripting of complex queries that combine free text and wild card searches.

A number of serious drawbacks in the ED dataset collected and distributed by SPARCS were identified that precluded it from being utilized in the analysis. Specifically, these drawbacks include:

1. Because individual patients cannot be identified in the de-identified SPARCS dataset[8], there is a high potential for counting the same patient more than once. This is especially problematic in the underserved areas where EDs may be over-utilized by the residents due to the lack of primary care options (Ford et al. 2008; Barron et al. 1999).
2. It is common that ICD-9 codes recorded in the ED during the visit include only those diagnoses that were the cause of the chief complaint and led to the ED visit. Since diabetes is a chronic and complex disease that rarely presents as the primary complaint during the typical ED visit, it may be underreported in the outpatient SPARCS database. The EMR alternative of diagnosing diabetes, for example by glycated hemoglobin test results, can be used.
3. In New York, the governmental agencies that license food establishments typically can only provide data for the establishments licensed within the current calendar year. In urban areas, however, the local food environment can change rapidly due to changes in economic conditions (e.g., re-zoning or availability of commercial space in the area), in the local population, and in market forces at large. For example, it has been estimated that 26.2 % of independently owned restaurants fail and close in their first year of operation (Parsa et al. 2005). Compounded with this is the fact that there is usually a one to two year gap between the request for data from the SPARCS IRB and the actual receipt of the health data. This creates an inevitable temporal mismatch between the health outcomes and food environment data, therefore introducing bias into the food environment data and making the accuracy of all food outlet data unverifiable.

[8] De-identified SPARCS was the only SPARCS dataset that could be obtained to meet the timeline for this study.

Initially, to measure the prevalence of diabetes in the study area, the use of the outpatient or ED dataset distributed by the SPARCS was considered. However, after a careful review of the relative strengths and weaknesses of the SPARCS data as compared to the data collected from an extensive electronic medical record (EMR) database, the latter was used. The decision to use the EMR database rather than the SPARCS dataset was based on the finding that the EMR database provided both the most accurate street address-level residence location data on the subjects and the most robust dataset to detect both new and existing diabetes cases.

The EMR-Derived Database

The individual-level health data was extracted from the Emergency Department (ED) visits at the Kings County Medical Center (KCMC), the largest hospital in Brooklyn, NY, serving central and northern Brooklyn, for the six month period between May 1, 2008 and October 31, 2008. The initial raw data contained 60,103 ED visit records. The raw data extract from the EMR oracle database contained patient's demographic information, anthropometric and medical data in ICD-9 codes format, patient's addresses, in order to permit precise geocoding, and patient's race/ethnicity information.

The use of the EMR data extracts as the source of all health data for the study necessitated the creation of a large separate study database with a complex data management schema. The original data (60,103 ED visits) was housed on an Oracle-based electronic medical record database at KCMC. The data was then transferred from the Oracle database, at the KCMC, into the study database in MS Access 2007. A summary of data elements requested to be extracted from the EMR are shown in Table 1. Unfortunately, some of the data elements listed in Table 1 were not available. For example, data on height and weight, used to calculate BMI, are rarely measured in the ED and were available only for 71 subjects (less than 5%) and thus was not used in the selection of cases and controls or the analysis. Therefore BMI could also not be included in the analysis.

After the import was completed the 60,103 ED visits were queried for repeat visits using a simple SQL query. The SQL query revealed that 30.9% of the 60,103 visits were return visits to the ED at least once during the six month study period. After subtracting the 18,546 duplicates from the analysis, it was concluded that there were 42,454 visitors to the ER during the study period (any repeat visits to the ED during the study period were recorded in the EMR). This extremely large number of repeated visits by the same individuals is consistent with reports by the Health and Hospitals Corporation or HHC, which manages this and other New York City owned hospitals. Due to KCHC's location in a medically underserved area, many individuals, especially those with chronic conditions and without medical insurance, utilize its ED for primary care needs. The 42,454 patients were given a unique 10-digit study identifier.

The Institutional Review Board (IRB) application for the 2008 patient data was approved and all Health Insurance Portability and Accountability Act (HIPAA) regulations were followed to protect the confidentially of the subjects in this study. The process of de-identifying the individual medical records will be discussed later in this chapter.

Geocoding of Subjects to Residences

One of the goals of this study was to test the feasibility of tethering health and address data stored on the EMR of patients who visited the KCMC ED to a GIS. The overriding rationale for such a linkage is that in the future such tethering of health events to a GIS—which contains a wide array of geo-referenced socio-economic, demographic, food, and built environmental data—may hold the potential to better understand the underlying pathways of metabolic and other diseases. Because the subject data came from the EMR, a relatively novel source of data for health research, one corollary product of this study is the evaluation of the feasibility of such tethering of a hospital EMR to the GIS.

After assignment of the unique study IDs, each patient's address was geocoded using ArcGIS 9.3 software from ESRI. For patients that visited the ED more than once during the study period, the address recorded during the last visit to the ED was assumed to be the current address and was used in geocoding of the patient addresses. After the initial geocoding run, the resulting shapefile was reviewed manually. As a result, 2,313 individuals with addresses outside of Brooklyn were removed and the additional 960 individuals with Brooklyn addresses that could not be geocoded to an exact street address were removed from the study database. The total number of geocoded patients with valid addresses in the study area was 39,201.

The study subjects resided in areas with diverse socio-economic, ethnic, and racial compositions, representing a wide cross-section of Brooklyn neighborhoods. The highest density of the study population was in East New York, Crown Heights, and Central Flatbush.

An explicit attempt was made to not only acquire the most accurate address data available on residences of the subjects of the study but to also protect subject's confidentiality. To protect confidentiality of the study subjects, a mechanism was developed through which, after geocoding the addresses, all address identifier information, except longitude and latitude, was deleted from the study database. For data display, the points representing the location of residence within the respective census blocks were randomly plotted. This approach spatially masked the locations of individual subject's residences within their census blocks to protect the subject's confidentiality while still permitting data pattern visualization. This process also permitted a linkage to the neighborhood SES variables at CT level, discussed later in this chapter.

1. Patient Name
2. MRN
3. Birthdate
4. Age
5. Sex
6. Gender
7. Census Block of home address
8. Health Insurance
9. Height
10. Weight
11. BMI
12. Admission Diagnosis
13. Previous Medical History
14. Glucose (mg/dl)
15. HGB A1C
16. ED disposition
17. Prescribed Medications

Table 1. Data elements requested to be extracted from the EMR.

Race/Ethnicity Data

Reliable race data was available only for patients who were listed as African-American or black. This important limitation of the data was an artifact of a poor data entry mechanism for the fields reflecting race and ethnicity of the participants in the ED. At KCMC, race and ethnicity information is entered in to the patient's EMR by the receptionist. There is only one field into which both race ethnicity information is entered. Clearly, such a conflation of these two distinct variables was problematic. Furthermore, the Race/Ethnicity field in the EMR contained a drop-down menu with only two options—"African-American" and "Black". If the individual's race was not identified as

African American or black, their race and ethnicity could be entered as a free-text into the same field by the ED receptionist. Such free-text entries were often confusing because of the many misspellings. In other instances these free-text entries were erroneous because did not reflect racial or ethnic status of the individual. For example, free-text entries included terms such as “Middle-Eastern” and “English”. Clearly, these terms did not reflect one’s race or ethnic origin because they only denoted the geographic origins of the particular patient. Ethnicity of the patients could not be used in the analysis because of such data entry errors. Due to the shortcomings of the source data in respect to race described here, there was no choice but to exclude all individuals whose race was not recorded properly and race was consistently and clearly recorded only for these individuals identified as African American or black. Since race has been shown to be an important factor in diabetes (Gwynn et al. 2011; Congdon 2010; Ettner et al. 2009; Auchincloss et al. 2008; Harris 1991) it was essential to limit the analysis to the group whose race was properly recorded. Therefore, only individuals identified as African American or black were chosen for participation in this study. After the exclusion of all individuals not identified as black or African American, the study sample was limited to 35,673 persons. Based on the methodology described in the next section, cases and controls were selected only from these individuals identified as African American or black.

Selection of Cases and Controls

Individuals’ medical records in the KCMC dataset were reviewed. All adult patients identified as black and African American who were treated in the ED of KCMC during the study period, and resided in Brooklyn were considered for the selection of cases and controls. The decision to include only the adult subjects was due primarily to the infeasibility of obtaining parental consents to include children as subjects in this retrospective study. The decision to include only the individuals whose race was identified as African-American or black was due to the limitations of the source data. Specifically, only race and ethnicity of the participants who were identified as African-American or black was consistently recorded in the EMR. This data limitation is also discussed in chapter 5.

According to the American Diabetes Association (ADA), the glycated hemoglobin test is a newer and more accurate test than the traditional blood glucose test in the measurement of a person's blood glucose level. It has been deemed the blood test of choice for the diagnosis of diabetes by the ADA because it measures the mean blood glucose level over the past 3 months making it less susceptible to short term fluctuations in blood glucose levels. Commonly called the HbA1c test, it shows the amount of glucose that sticks to the wall of a red blood cell, which is proportional to the amount of glucose in the blood (http://www.diabetes.org/diabetes-basics/common-terms/, accessed on 6/19/2011). Normal levels of glucose in the blood produce a low level of HbA1c, however, in diabetes, as the average amount of plasma glucose increases, the proportion of HbA1c also grows. The American Diabetes Association and the European Association for the Study of Diabetes now recommend the use of the HbA1c test for diabetes mellitus diagnosis if the HbA1c level is greater or equal to 6.5% (Gomez-Perez et al. 2010).

Subjects were categorized as follows:

Diabetes Cases (n=3,202)

A. Included all codes pertaining to current or past diagnosis of diabetes mellitus (excluding gestational diabetes diagnosis codes) (n=2,389)

The diagnosis and ICD-9 codes are as follows:

1. Diabetes Mellitus:

    a. Type 1 (ICD-9 code: 250.01)
    b. Type 2 (ICD-9 code: 250.02)

2. Diabetes Mellitus related physical health conditions:

    a. Diabetes with renal manifestations (kidney disease) 250.4
    b. Diabetes with neurological manifestations 250.6
    c. Diabetes with other specified manifestations and Diabetes with other unspecified complications 250.8 and 250.9, respectively

B. All Subjects with no prior history of diabetes but with elevated HbAlc at or above 6.5% at the time of their visit (n=813).

Control Group (n=7,518)

Specific steps were then taken to identify an appropriate control group. Subjects who did not meet the Case Group selection of ICD-9 codes for diabetes, prior history of diabetes, and an HbAlc level below 6.0 % were considered for the Control Group. Specifically, persons free of diabetes and diabetes related illnesses were selected using the list of diagnostic codes that excluded nutrition and metabolism-related diseases. Importantly, all individuals that did not meet criteria A or B of the cases group but had blood glucose levels above the normal limits, regardless of their diagnostic codes, were also excluded from the control group. After the initial removal of any subjects that did not meet these basic selection criteria, the controls were identified using a list of diagnoses that are not associated with metabolic balance or nutrition. Individuals with metabolic and diet-related conditions, such as these with cardio-vascular disease, renal disease, and stroke were excluded from the control group. The final selection yielded 7,518 individuals for the control group.

Definition of the Neighborhood Boundaries

The study spatially linked individual-level health data, extracted from patient's records, to their individual food environments. In order to accomplish this task, a robust GIS dataset that encompassed neighborhood food and SES environment data was constructed. Data sources for these variables will be discussed in some detail later. At the onset of the study it was unknown to what degree residents with and without diabetes differed in their respective food and socio-economic environments in Brooklyn. Because this information did not exist, it was impossible to measure whether specific local food and socio-economic environments have protective or detrimental effects on the development of diabetes and other health outcomes. Therefore, measuring detailed characteristics of neighborhoods around residences (neighborhoods were proxied by CTs) will provide important insights into this question for future studies. Specifically, the food environments were assessed by measuring a patient's proximity to healthy and unhealthy food options and the density of these outlets. These food options were assessed through the coding of food service establishments (e.g., restaurants, cafeterias, etc.) and food stores (e.g., supermarkets, bodegas, etc.). The goal was to highlight the neighborhoods in which a variety of fresh and prepared foods, including fruits and vegetables can be obtained versus the neighborhoods dominated with stores and food service establishments that primarily sell unhealthy foods.

To analyze how variations in the context affect the outcome of diabetes, the study area needed to encompass a wide range of neighborhoods that provide different types of food and socio-economic contexts for their residents. Brooklyn, because of its unique diversity, was an ideal choice. The initial mapping of the spatial distribution of diabetes cases in 2008 located large numbers of patients with diabetes in three UHF neighborhoods in Central and Eastern Brooklyn. Based on the NYC HANES data, all three of these UHF neighborhoods have the highest rates of diabetes in Brooklyn and some of the highest obesity rates in New York City (Figure 1).

The central challenge posed in this study was to detect which variables within the built environment may engender diabetes and other metabolic spectrum diseases. A high degree of accuracy in geocoding individuals to their addresses was especially important here because the ability to explain the effect of differential access to healthy and unhealthy food outlets on the outcome of interest depended on the accuracy of the geocoding. Such precision is critical to test if small area, block-by-block differences in access to various foods *near* residences have an effect on the outcome of diabetes, while accounting for the potential neighborhood-level SES confounders, such as income and education of the residents.

The task of selecting the appropriate neighborhood boundaries was key to this study. Selecting neighborhood units that are too large could "smooth" the sharp differences in SES that often exist in adjacent neighborhoods. After all of the cases and controls were selected and mapped, it became evident that the study sample lives in 62% of the census tracts in Brooklyn. The smallest Census tabulation area to which socio-economic data can be aggregated is block group, but many block groups have relatively small populations. Therefore, using the block groups to define the SES of the neighborhood would have reduced the number of individuals for whom SES status was imputed, thereby introducing larger standard error to the analysis. Thus, for the purpose of this study I defined the individual neighborhoods as U.S. Census 2000 Tracts. The aggregated Census 2000 SES data on the CT level was later joined to the points that represented geocoded residences of cases and controls to account for the neighborhood SES effect on the outcome. The study subjects lived in 452 out of the total of 782 Census tracts in Brooklyn. On average, a CT in the study in which the study subjects lived, housed 20 subjects (SD 29.204), with the highest number of subjects per CT being 229.

Neighborhood Environment Data

It is currently unknown to what degree residents with and without diabetes differ in their respective food and socio-economic environments. Because this information does not exist, it is unknown whether specific local food and socio-economic environments have protective or detrimental effects on the development of diabetes. Measuring detailed characteristics of the unique block-by-block foodsheds around residences can provide important insights into this question. Specifically, a major goal of this study was to measure the association of the local food environments with health outcome. These spatial access measures were constructed based on walking times to reach various healthy and unhealthy food options. The individual's total food environment was constructed through the coding of food service establishments (e.g., restaurants, cafeterias, etc.) and food stores (e.g., supermarkets, bodegas, etc.). The goal was to characterize neighborhoods by access to a variety of fresh and prepared foods, including fruits and vegetables, then determine which neighborhoods are dominated by stores and food service establishments that primarily sell unhealthy foods, and finally, measure this neighborhood effect on diabetes within the sample of local residents.

Food outlets data

For this study, two important types of food outlets, were considered food stores, which primarily sell food items, and food service establishments, which serve prepared food for eat-in or take out. In order to adequately characterize the total food environment, the NYC Department of Health and Mental Health and the NY State Department of Markets and Agriculture datasets, listed in Table 2, were utilized.

Since different state and local government bodies regulate food stores and food service outlets, several complementary sources were utilized. In New York City, the New York Department of Health and Mental Hygiene has a mandate to annually inspect and license qualifying food service establishments, while the New York State Department of Markets and Agriculture annually inspects food stores and food markets across New York State. Through several Freedom of Information Act (FOIA) requests, detailed data records on all Brooklyn food service establishments were obtained, including all operating restaurants, cafeterias, and sandwich shops inspected by the Departments of Health and Mental Hygiene in 2008. Also using FOIA, a 2008 food store inspections dataset compiled by the NY State Department of Markets and Agriculture was obtained. Both datasets include address, longitude and latitude data, business name, address and ownership information, square footage, total number of employees, and number of checkout cashiers for each store. This extensive and detailed food environment dataset was reformatted and coded to be utilized extensively to study the relationships between diabetes and the local food environments that were hypothesized to engender diabetes.

Data Source	**Data Type**	**Method used to locate the outlets**	**Variables recorded in the data source**	**Match of outlets in the electronic data source and field survey**
New York City Department of Health and Mental Hygiene	Restaurant Inspection Records (RIR)	Geocoded by the author using restaurants' street addresses	Restaurant name Doing business as name (DBA) Type of service (e.g., waiter, takeout, limited seating) License number Address Cuisine Code Cuisine Code description	92% match between geocoded restaurants' and field survey

NY State Department of Markets and Agriculture	Food stores licensed by NY State	Department of Markets and Agriculture provided longitude and latitude of each food store	Store Name Doing business as name (DBA) Number of employees Square footage Number of check-out registers Address	100% match between food stores in the electronic data source and field survey
Food stores: ESRI Business Analyst	All food retail establishments Both	MapUSA provided geocoded food stores and restaurants	North American Industry Classification System Codes (NAICS) Store name Doing business as name (DBA) Number of employees Square footage Number of Check-out Registers	62% match between food stores in the electronic data source and field survey, many small outlets physically present during field survey were missing in MapUSA database

Table 2. Sources of food outlets data.

InfoUSA, a commercial business listings dataset was considered but found to be unreliable (Table 2). The InfoUSA Business Analyst is available through the Business Analyst extension to ArcGIS software (Redlands, CA).

InfoUSA Business Analyst is a commercial product that provides a national database that has approximately 12 million U.S. businesses arranged by business name, industry description, North American Industry Classification System (NAICS) code, sales volume, number of employees, square footage, and address location. However, after mapping the address data for a sample census tract with a high frequency of subjects[9] and comparing the InfoUSA business location data for food stores with the location data obtained

[9] Census tract with over 200 individuals in the sample lived in that census block.

through a field survey, only 62% of all listings in the InfoUSA dataset could be matched to the census block numbers obtained by the field survey conducted by the author in the winter of 201. Three CT's in Brooklyn were randomly selected for the survey and surveyed by driving along all street segments and visually scanning all street faces for food outlets and recording businesses names and their geographic coordinates using a GPS-enabled portable devise. These records were then compared against the three databases listed in Table 2. The results are reported in the last column of the table. Consecutive accuracy tests showed large discrepancies between InfoUSA's store locations and business types, and the ground-truthed dataset[10] obtained through the field survey.

While NAICS codes provided a useful starting point for the classification, a review of the NAICS documentation and salient literature on the food environment made it clear that a more fine-grain coding of outlets was necessary. Therefore, the food outlets were assigned a numeric rating, from healthy to unhealthy, which was used to put the food outlets into distinct categories. The decision was based on the overarching assumption that outlets selling a variety of food choices, even if not all of them were healthy, provide healthier diet options than those outlets which provide a very small number of food options, usually of the fast-food type. Additionally, one category was assigned to food outlets whose classification could not be readily determined.

---

[10] NAICS codes for food outlets # 445110, 445120, 447110, 445210, 445220, 445230, 445291, and 445292 were used for comparing food outlet types and locations in the databases produced by the NYC DOH, NYS Dept. of Markets, and InfoUSA to the benchmark field survey dataset for accuracy. Field survey data was considered accurate at the time of the comparison.

Initially, groupings of stores and food service establishments using the NAICS industry codes[11] were broadly defined. However, these codes were too broad to capture the nuances of the local variations in the food environment in the neighborhoods where the subjects lived. For example, NAICS codes do not distinguish between supermarkets and smaller grocery stores (NAICS code 445110). Previous research on the food environment indicated that such a distinction is necessary to accurately characterize the food environment in urban areas. Specifically, Morland et al. (2002) showed that supermarkets typically have more healthy food items at lower costs, thus highlighting the importance of distinguishing them from smaller grocery stores. Other researchers showed that supermarkets typically provide a greater variety of fresh fruits and vegetables, offer fresh meat and fish, and have more choice in other food groups, including a wider choice of dairy products, whereas smaller grocery stores usually have less choice and sell food at higher prices (Galvez et al. 2008). Using the methodology based on the food outlet coding developed by Morland et al, a coding schema for a more precise scoring of the food outlets necessary for this small area study was created (Morland and Filomena 2007). The overall strategy was to separate outlets that offer the least choice among foods and serve mostly fast-food from the outlets that sell or serve a wider variety of food choices, including diabetes-healthy food options (Karnik et al. 2011; Horowitz et al. 2004). The former category thus included bodegas, convenience stores, delis, and pizza parlors. The latter category included supermarkets, fruits and vegetables markets, restaurants that offer a variety of foods presumed to be diabetes-

[11] Detailed information on the North American Industry Classification System can be found on OSHA's web site http://www.osha.gov/oshstats/naics-manual.html.

healthy. Corporate-owned chain supermarkets (e.g., Price Chopper, PathMark, Whole Foods) were also separated from the smaller, independently-owned grocery stores (e.g., Gerry's Grocery, West Side Market) because they were reported to substantially differ both in choice of foods and in price (Morland and Evenson 2009, Morland et al. 2002). In order to distinguish between the many types of food outlets, a SQL script that permitted the assignment of more precise codes for all food outlets was used. This approach is reflected in the food environment classification listed in Tables 3 and 4.

Table 3: Food Store Classification

Establishment type	**Study codes**	**Study code definition**	**Examples**	**NAICS code**	**Abridged NAICS definition**
Meat markets, Fish markets	45	Butchers and meat markets, seafood stores	Lewis Meat Market, Kim's Fish Market	445210, 445220	Establishments primarily engaged in retailing fresh, frozen or cured meats, poultry or fish
Fruit and vegetable stores	6	Stores that sell primarily fruits and vegetables	Song's Fruit & Vegetable Store , Utica Fruit Market	445230	Establishments primarily engaged in retailing fresh fruits and vegetables.

Bakeries, baked goods, confectionery and nut stores	78	Bakeries, confectioners, and nut stores	Hammond's Finger Lickin' Bakery Honeybee Candy Store	445291, 445292	Establishments primarily engaged in retailing baked goods
All other food stores	99	All other specialty food stores	Bklyn Larder, Bierkraft	44529912	Establishments primarily engaged in retailing miscellaneous specialty foods not for immediate consumption

Table 3. Sources of food Store Classification

[12] The food stores in this category were removed from the analysis based, on Morland, 2007, because their contribution to the total food environment was assumed negligible.

Table 4: Food Service Establishments					
Establishment type	**Study codes**	**Study code definition**	**Examples**	**NAICS Code**	**Abridged NAICS definition**
Cafeterias and food buffets	60	Establishments providing a wide choice of foods, including fruits vegetables. Patrons select from food and drink items on display in a continuous cafeteria or buffet line	Home Town Buffet, Polish Slavic Center Cafeteria	722212	Establishments primarily engaged in preparing and serving meals using steam tables, a refrigerated area, and self-service nonalcoholic beverage dispensing equipment
Full service restaurants	60	Restaurants providing a wide choice of foods, including vegetables	Paradise Garden Restaurant, Morton's the Steakhouse, Chez Lola	722110	Restaurants, full service, steak houses, full service, fine dining restaurants

Table 4: Food Service Establishments (continued from previous page)					
Carryout fast-food	70	Franchised and non-franchised fast-food outlets	McDonald's, White Castle	722211	Fast-food restaurants, Pizza parlors, limited service restraints, pizza delivery shops
Carryout sandwich shops	70	Delicatessens and sandwich shops	Corner Deli, A&R Deli, Grab-n-Go Deli	722211	Delicatessens, sandwich shops
Bagel shops	70	Bagel shops	Tasty Bagels, La Bagel Delight	722110	Bagel shops, full service
Carryout snack and nonalcoholic beverage bars	80	Carryout specialty items	Starbucks, Joe's Coffee Bar	722213	Beverage (e.g., coffee) bars (non-alcoholic), Doughnut shops, ice cream parlors, pretzel shops
Bar and taverns	90	Drinking establishments	Last Exit Bar & Lounge, Gutter Pub	722410	Alcoholic beverage drinking places

Table 4. Food Service Establishments

For analytical purposes, all of the outlets coded into a particular outlet category were assumed to provide access to the same types of foods (see column 2 in Tables 3 and 4). Therefore, a between-category difference was assumed to be of key importance in defining the local food environment. For example, for analytical purposes it was assumed that there are no differences between cafeterias (NAICS code 722212) and full-service restaurants (NAICS code 722110). Importantly, it would be an oversimplification to state that access to healthy food is defined by spatial access to outlets alone. Explicitly, in addition to the effect of walking access to food outlets (expressed by walking time to stores, and as the density of a particular store type per subject's neighborhood), the food price and nutritional quality of the foods sold inside the outlet should also be assessed. Such an "in-outlet" assessment is important because food quality, choice, and prices inside the outlets (stores or restaurants) can be presumed to have an effect on diets and thus, albeit indirectly, on the health outcome of interest, diabetes.

For example, if a grocery store has a produce stand but the produce is of poor quality (old, unattractive, or withered) and there are few choices available among the produce items, shoppers may choose to avoid buying produce altogether and purchase processed, fast-food type items instead, in effect excluding fresh produce from the diet. Similarly, if the price for the specific healthy food items is higher than less healthy options, it may pose an economic barrier to incorporating these healthy food choices in the diet. Both the former and latter scenario are frequently cited as barriers reported by the diabetes patients interviewed, especially those who live in the underserved neighborhoods, where the overwhelming majority of the subjects in the present study sample reside (personal correspondence with Dr. Mary Ann Banerji, Diabetes Clinic Medical Director, Kings County Medical Center).

Neither InfoUSA nor the NYS Department of Markets and Agriculture datasets provided a way to measure the in-store food environment. The information provided allowed for characterization by store name, the doing business as (DBA) name, NAICS code, square footage, and number of employees. Based on the information given it was not feasible to sample and analyze the in-store food environments for a more precise understanding of stores within the same type (see column 2, Table 3) due to the time and field work necessary to complete such a task. Data on food service establishments were somewhat more complete than store data in terms of the descriptive characteristics of the food served in-outlet.

The field sampling of in-outlet food cost and nutritional quality was also deemed impractical and other research that addressed this issue was relied upon (Kim et al. 2010; Morland and Filomena 2007). The dataset provided by the New York City Department of Health and Mental Hygiene, which consisted of annual inspection records, proved resourceful for information on the in store food environment. In New York City, restaurants are typically scheduled for an unannounced NYC Health Department inspection at least once a year. During the inspection, an inspector records the type of food service and cuisine served, checks for compliance with city and state food safety regulations and assigns points for any condition that violates these rules. The information contained in these NYC Health Department inspection records was used to assign specific codes to all the licensed food outlets in Brooklyn. These annual inspection records included data on the type of food service (e.g., fine dining) cuisine served (e.g., pizza) and the square footage of the retail space of each establishment using SQL queries. Similarly, SQL queries were used to assign an outlet type to all food stores. These food outlet codes were then used in the regression models reported in Chapter 4.

A total of 5,313 food service establishments and 4,426 food stores in Brooklyn were identified, geocoded, and assigned food outlet type codes described in Tables 3 and 4, utilizing the SQL scripts and manual record by record review. A major limitation of the food environment dataset, based on the governmental inspections used in this model, is that they provide virtually no information on the cost or nutritional quality of the foods sold in a particular store or restaurant.

Neighborhood SES Data

Previously published research has demonstrated that low neighborhood SES has been linked to obesity, heart disease, and higher mortality rates among its residents as compared to wealthier neighborhoods in New York City and other U.S. cities (Darmon 2008). It is reasonable to hypothesize that in poorer areas of a city, the food environment could be degraded and thus would provide fewer healthy food choices to the local residents. This claim has been advanced by several recent studies (Curtis 2010; Galvez et al. 2008; Lopez-Zetina, Lee, and Friis 2006). According to some researchers, the SES disparities in some parts of New York City narrowed somewhat between 1990 and 2000 (these trends may have been influenced by the in-migration of wealthier populations and gentrification in some neighborhoods), while SES disparities, including poverty and low educational attainment, remained unchanged in other parts of the city (Althoff et al. 2009). Throughout this study it has been presumed that the neighborhood's socio-economic conditions may influence spatial and economic access to the food outlets and thus influence diets due to available food choices. Such influences may, in turn, affect metabolism and indirectly influence the risk of diabetes. Therefore, several neighborhood-level SES variables were created to measure the effect of the local socio-economic environment on the outcome of interest. The source data was downloaded from the Census 2000 Summary File 3[13] and expressed as percentages of the CT population. The methods of imputation and the accuracy of the data are documented in the Supplementary Survey Summary Tables (Summary File 3 Documentation, U.S. Census 2000).

Methodology for Computing Neighborhood SES and Educational Attainment Markers

In order to calculate the indices of the neighborhood SES, the SES calculation methodology developed by the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) was used (http://seer.cancer.gov/seerstat/variables/countyattribs/ , accessed on 12/1/2010). The SEER scripts were modified to meet the needs of this study. SEER is an authoritative source of methodology used for computing SES and was developed to track the incidence of cancer and survival rates in the U.S. The SEER databases currently collect and publish

[13] Census 2000 Summary File 3 (SF 3) - Sample Data, http://factfinder.census.gov/

cancer incidence and survival data from state-run, population-based cancer registries covering approximately 26% of the United States population (Harper and Lynch 2007). The SEER Program registries routinely collect data on the SES of the area where the patient resides as well as the individual's demographics, primary tumor site, tumor morphology, stage at diagnosis, first course of treatment, and follow-up for vital status. Currently, the SEER database is the only national disease registry in the U.S. The methodology used for the SES data extraction from the US Census SF1 and SF3 2000 databases (http://www.census.gov/main/www/cen2000.html) and for the SQL scripts, were adopted from the SEER methodology (Harper and Lynch 2007; Weiss 1988).

The process of computing the subject's SES was conducted as follows. The subject's residence was spatially joined to the neighborhood of residence (with the census tract of residence defining the boundaries of the individual's neighborhood). CTs with a total population of less than 100 persons were excluded from the neighborhood SES calculations because the population denominator was considered to be too low to reliably measure the aggregated neighborhood SES. Five CTs that had population counts of less than 100 persons were thus excluded from the SES calculations.

Three educational attainment markers were calculated from the Census SF3 table P37 (Sex by Educational Attainment for the Population 25 Years and Over). The three education variables included were: (1) less than 9th grade education; (2) less than high school graduate; and (3)

The Following SEER adopted SQL scripts were used
(http://seer.cancer.gov/seerstat/variables/countyattribs/):

1. Less than 9th grade (EDU_LESS9G):

   ((P037003+…+P037006+P037020+…+P037023)/P037001)*100
2. Less than HS:

   ((P037003+…+P037010+P037020+…+P037027)/P037001)*100
3. HS Diploma: ((P037003+…+P037010+P037020+…+P037027)/P037001)*100

Measure of Poverty

The SEER methodology was used to categorize people and their household members as above or below the poverty line and also to measure the degree or depth of poverty based on the ratio of income to poverty in each CT (Harper and Lynch 2007). The ratio of income to poverty compares an individual's income with their poverty threshold, and expresses that comparison as a fraction. Ratios of 1.00 or greater (100% of poverty or greater) indicate income at or above the poverty level, whereas ratios below 1.00 (below 100% of poverty) indicate that the individuals income is below the federal poverty level. People with incomes at or above the threshold but below 1.25 (125% above poverty) of the threshold are sometimes classified as "near poor." The Census Bureau describes those individuals living in the households with incomes below one half of the poverty threshold as "severely poor." For example, a poverty ratio of 1.0 (income at 100% of poverty level) means a person is living at the federal poverty line; a ratio of 0.5 (income at 50% of poverty level) would mean that the person is living in a household making only half of the income designated as the poverty threshold or "severely poor." The percentage of persons whose incomes are below the poverty level was derived from census SF3 table P88 (Ratio of Income in 1999 to Poverty Level).

The Following SEER adopted SQL scripts were used (http://seer.cancer.gov/seerstat/variables/countyattribs/):

1. Persons at and below Poverty Level: ((P088002+…+P088006)/P088001)*100
2. Severe Poverty, below 0.50 income/poverty ratio: ((P088002)/P088001)*100

Measure of Private Vehicle Ownership

It was assumed that access to a private car is an important factor influencing spatial access to food (Weiss 1988), namely because access to a car may be especially influential in food shopping patterns. The U.S. Census does not collect data on private vehicle ownership by individuals or families, but rather on car ownership levels per household. Also based on SEER methodology, the percentage of housing units without a car were calculated based on data from Table H44 (Vehicles Available by Tenure)

The Following SEER adopted SQL script was used (http://seer.cancer.gov/seerstat/variables/countyattribs/):

((H044003 + H044010)/H044001)*100

Percentage of Female-Headed Households with Children under 18 Years old

In order to calculate the percentage of single mother households table P100 (Household Type by Presence of Own Children under 18) was used.

The Following SEER adopted SQL script was used (http://seer.cancer.gov/seerstat/variables/countyattribs/):

((P010015)/P01001)*100

Percentage of Black Residents[14]

The percentage of black residents was expressed as the percentage of the total residents of the Census Tract (Morland, 2007). The Following SEER adopted SQL script was used (http://seer.cancer.gov/seerstat/variables/countyattribs/):

```
((P007003)/H007001)*100
```

Migration

Table P24 (Residence in 1995 for the Population 5 Years and Over) of the Census SF3 data was used to create five migration variables. These variables included percentage of persons: remaining in the same house (no migration); moved, but in the same county; moved, to a different county but in the same state; moved, to a different state in the U.S. and; moved, to outside the U.S.

The Following SEER adopted SQL scripts were used (http://seer.cancer.gov/seerstat/variables/countyattribs/):

1. Moved, same county: (P024005/P024001)*100
2. Moved, different county, same state: (P024007/P024001)*100
3. Moved, different state: (P024008/P024001)*100
4. Moved, outside the US: ((P024013+P024016)/P024001)*100

Measure of Foreign Born Residents

The percentage of persons who are foreign born was calculated using the Census SF3 table P21 (Place of Birth by Citizenship Status).

[14] Variables representing percentages of black residents, migration, foreign-born residents, and linguistic isolation were computed for mapping purposes only (not shown here) and were not used in the regression models.

The Following SEER adopted SQL script was used
(http://seer.cancer.gov/seerstat/variables/countyattribs/):

```
(P021013/P021001)*100
```

Linguistic Isolation Markers

The percentage of households that are linguistically isolated was calculated using the Census SF3 table P20 (Households Language by Linguistic Isolation). The Census Bureau defines linguistically isolated as a household in which all members 14 years old and over speak a non-English language and also speak English less than “very well” (have difficulty with English).

The Following SEER adopted SQL script was used
(http://seer.cancer.gov/seerstat/variables/countyattribs/):

```
((P0200004+P020007+P020010+P020013)/P020001)*100
```

Chapter Summary

This chapter provided an overview of the data acquisition, cleaning, and database management strategy. The chapter also provided an overview of the steps taken to clean the raw health, SES and food environment data and the coding approaches used. This labor-intensive process culminated in the creation of a fully geo-referenced dataset that was ready to be imported into a statistical software package to fit a set of regression models in order to explore the associations between the neighborhood SES, food access and the outcome of diabetes. This analysis and its findings are discussed in the two following chapters.

Chapter 4: Results

Introduction and Statistical Analysis Methods

So far the discussion has centered mainly on the acquisition, cleaning, visualization, and exploration of spatial and individual health outcomes data related to spatial patterns of diabetes and its determinates in the urban environment. In this section the cluster detection and regression techniques used to quantify the effects of the SES environment and food access variables on the risk of developing diabetes in the KCMC ED visitor will be introduced.

Firstly, the socioeconomic status of the neighborhood needed to be assessed (Rutt & Coleman 2005). Secondly, the food outlets in the individual's immediate neighborhood (Booth et al. 2005) and spatial access to these outlets needed to be included in the analysis (Handy et al. 2002). These relationships are important in the context of this study because each variable may have a different influence on energy balance and on glycemic control due to the effect on the diet. After the data was processed in the preparation for analysis, as described in the previous chapter, descriptive statistics on the proximity to different types of food outlets and on the SES status of the case and control groups were calculated. A spatial cluster detection technique was used to search for diabetes clusters within the study area. Lastly a multi level logistic analysis was performed. Logistic regression using a hierarchical linear model was appropriate for the analysis of this hierarchically organized data in which influences on binary health outcome of diabetes are hypothesized to come from more than one "level" of influence

(Congdon 2010). This data structure is often the case in neighborhood health research. HLM permits the inclusion of random effects influencing the outcome of interest (diabetes) that come from both individual level factors as well as neighborhood level factors. Individual level factors demographic and racial variables that have been shown in prior research to have an influence on the outcome of diabetes (Gwynn et al. 2011; Bader et al. 2010; Maroko et al. 2009; Auchincloss et al. 2008; Jamieson 2007). In addition to individual variables, the model should be fitted to estimate neighborhood effects and should include higher level effects coming from neighborhood level factors. The neighborhood level factors that were hypothesized to be related to diabetes were socio-economic contexts of each neighborhood as well as their respective food environments. Unlike traditional "one-level" logistic regression models, HLM protects the researcher from committing an ecological fallacy, where a spurious relationship may be found because the individual subjects in the model are attributed the mean socio-economic characteristics of the neighborhood where they reside. In contrast, HLM permits one to turn the traditional modeling technique around and ask the question of how the neighborhood context is influencing individuals in the sample. Random effects reported by the HLM permit us to estimate both the statistical significance and size of the neighborhood effect. For example, in some smaller neighborhoods, the size of the effect may be high, but because of the smaller number of observations, the significance level may be less and therefore great importance cannot be attributed to these smaller neighborhoods. In this sample, the average neighborhood, defined as CT, housed 20 subjects (SD 29.204). Variables were deemed significant at the $\alpha = 0.01$ level.

SES Factor Analysis

The socio-economic status of the neighborhood was calculated by creating a socio-economic deprivation index. A factor analysis was performed that yielded the four coefficients for each one of the SES markers included in factor analysis. The results of the SES factor analysis are presented in Table 5.

Neighborhood SES Marker	**Factor analysis results**[15]
Percent of individuals with less than high school diploma	0.540
Percent of households in poverty	0.997
Percent of housing units without a private vehicle	0.754
Percent of single mothers with children under 18 years of age	0.736
SS loadings	2.396
Proportion Variance	0.599

Table 5. Factor analysis results.

The data on each SES marker included in the factor analysis was derived from the CT-level 2000 Census data. Figure 4 shows the spatial distribution of study subjects in the study area. Figure 5 shows the percentage of black residents by CT in Brooklyn. Figures 6 and 7 respectively show percent of individuals with less than high school education, percent of households in poverty. Figure 8 shows the spatial distribution of CTs with the combined SES score, in which higher values indicate a higher degree of the overall socio-economic deprivation in the neighborhood of residence of the subject.

[15] Cronbach's Alpha was 0.789

Figure 4. Spatial distribution of study objects.

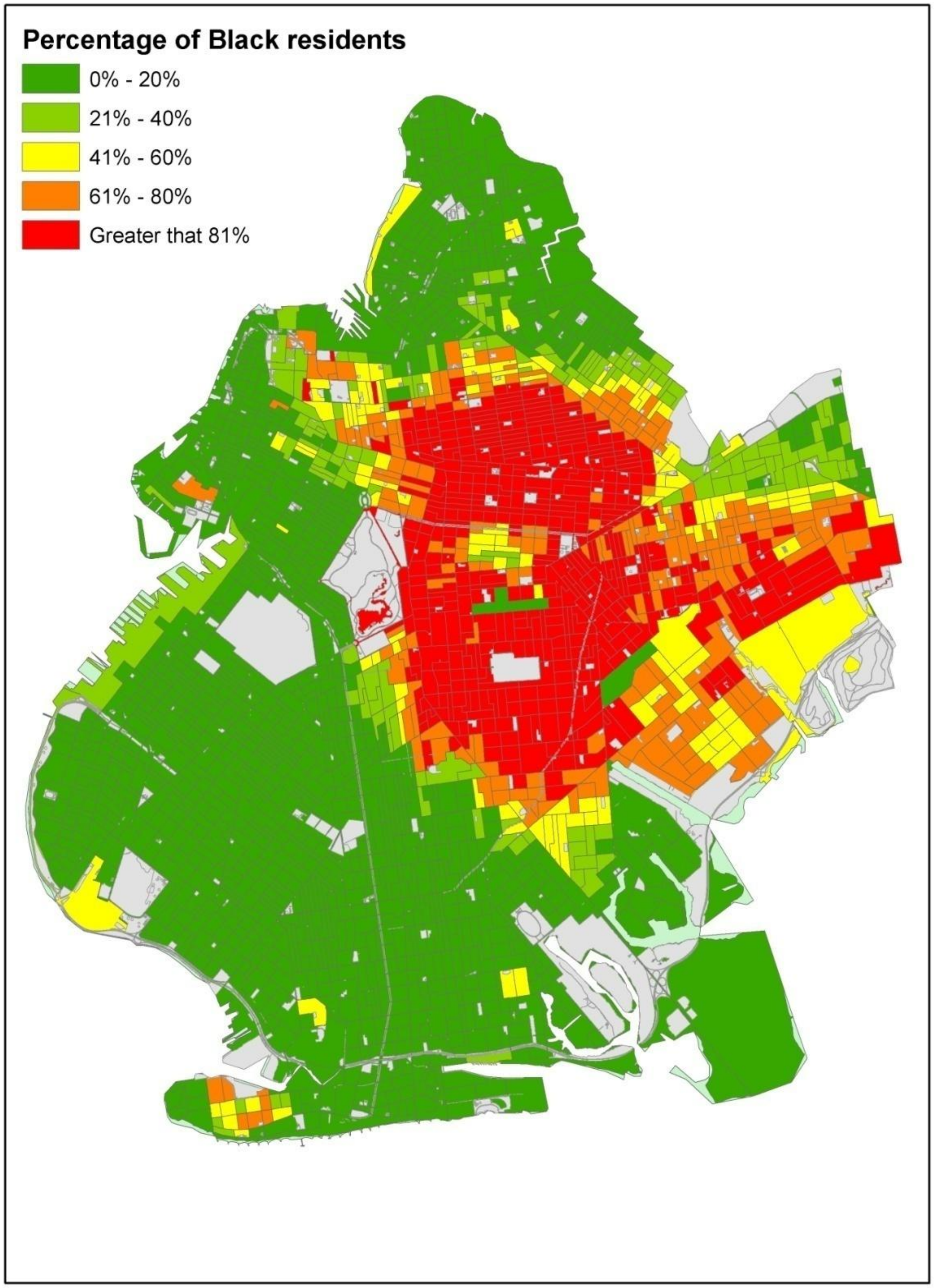

Figure 5. Spatial distribution of black population.

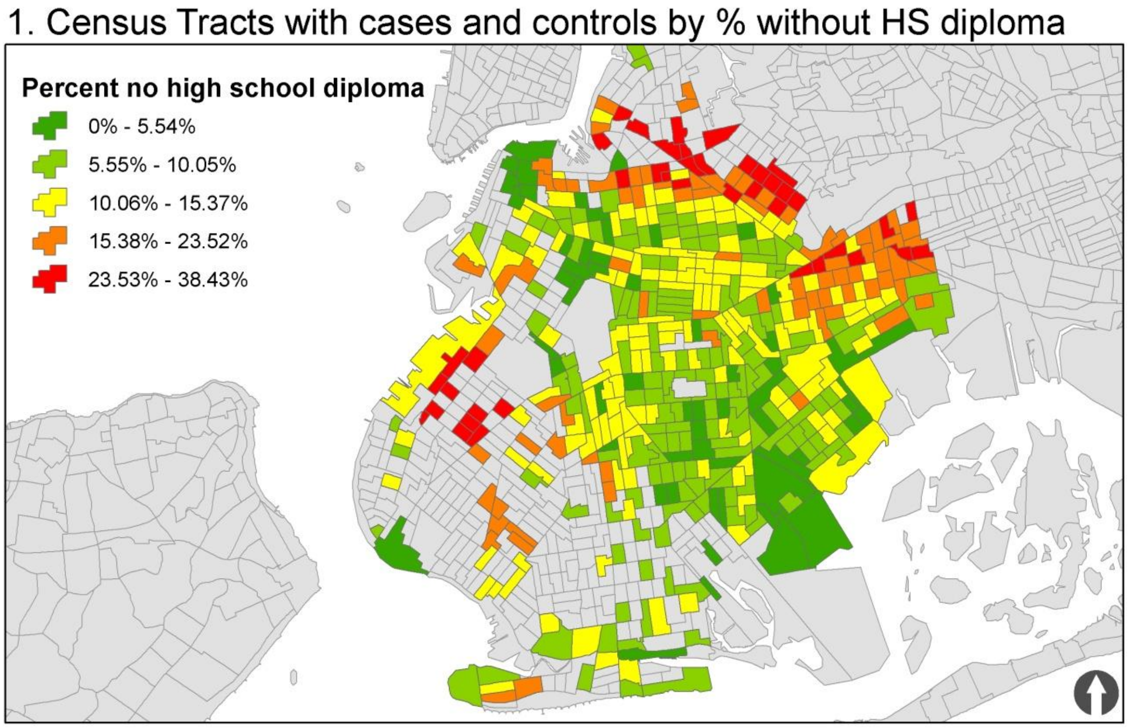

Figure 6. Residents without a High School Diploma

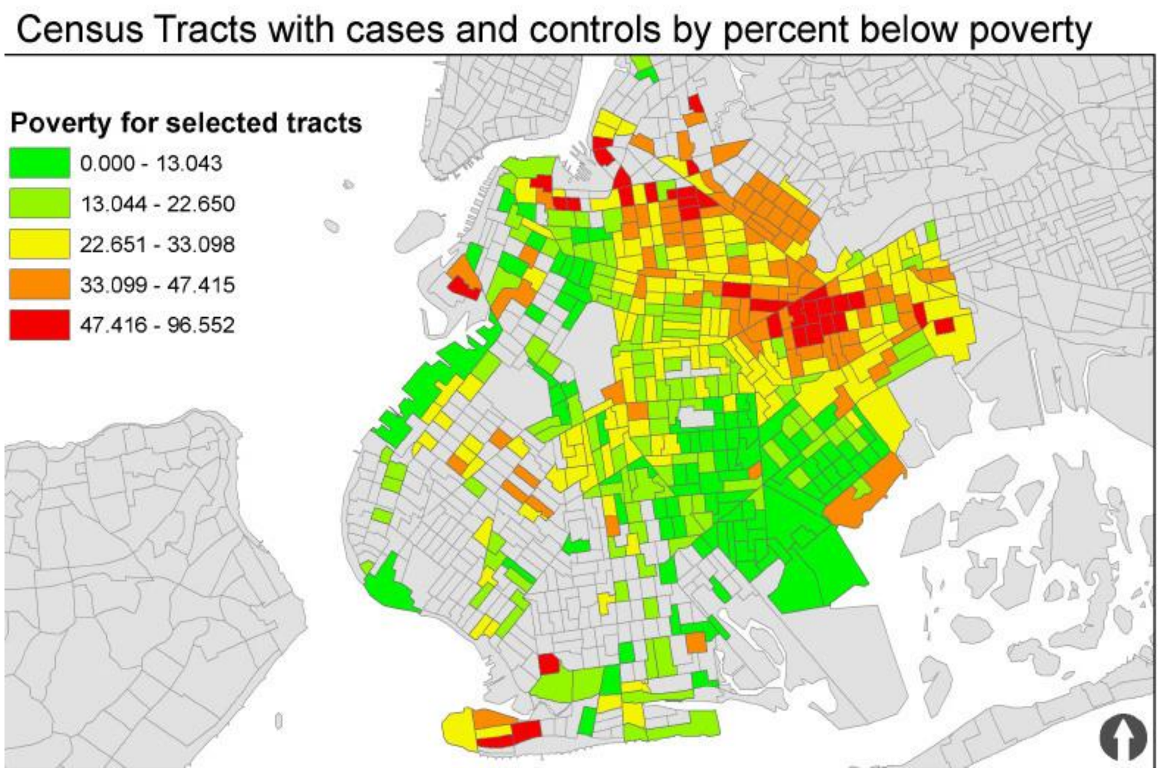

Figure 7. Percent of households in poverty.

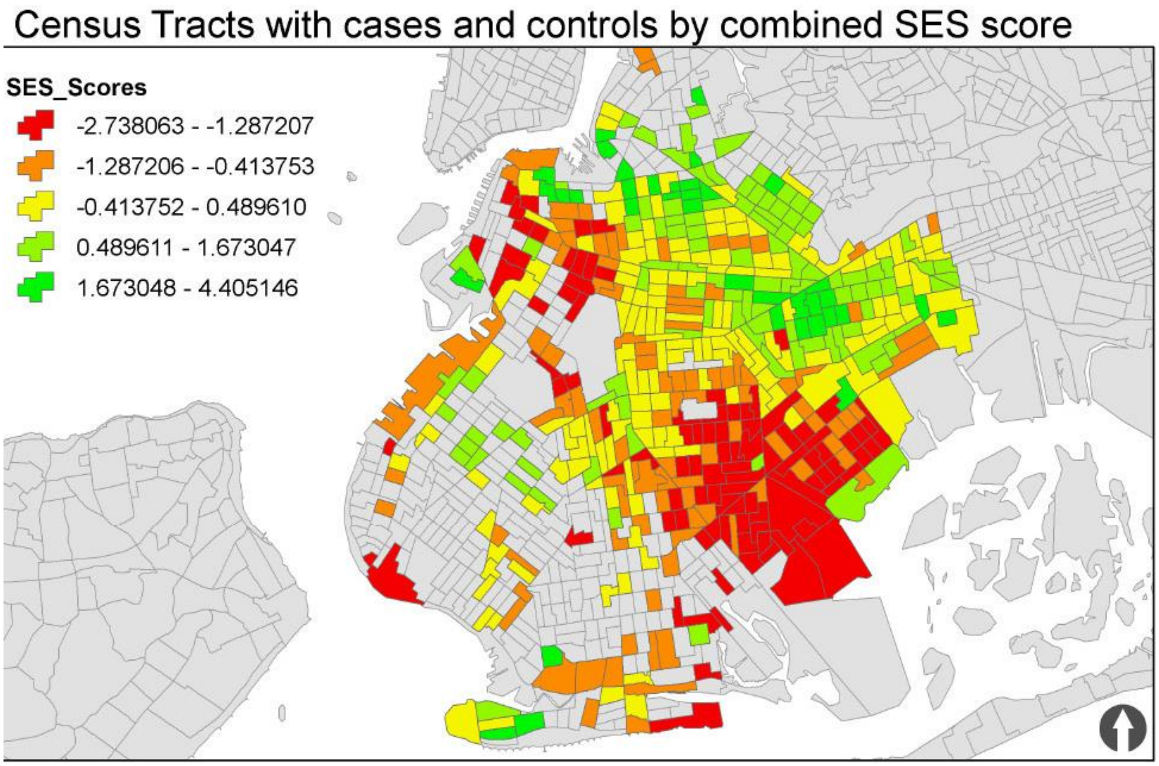

Figure 8. Composite SES deprivation score map.

The factor analysis produced a socio-economic deprivation score, with higher scores indicating a higher level of deprivation. Neighborhood composite deprivation scores showed that CT's with low single socio-economic indicator typically receive high composite SES deprivation scores.

Factor Analysis of Food Outlet Density

A factor analysis was also performed on food outlet densities in order to estimate the variance in food environments among the neighborhoods where the subjects lived. The density per census tract of food outlets, that was assumed to provide distinctively different food access in the neighborhoods of residence, was included in the food density factor analysis Table 6 provides summary of the results of the factor analysis for food outlet density.

Food outlet density per CT factor analysis results[16]	
Food Outlet Density per CT	***Factor***
Supermarkets	0.427
Grocery stores	0.717
Fruits and vegetates market	0.575
Bodega and deli	0.374
Fine dining establishments	0.362
Fast food establishments	0.831
Carryout snack and nonalcoholic beverage bars	0.482
SS loadings	2.222
Proportion Variance	0.317

Table 6.Results of the factor analysis on food outlets.

The Relationship between Diabetes and the Neighborhood Food Context

Utilizing the data on individual subjects and their neighborhoods, as well as insights from the salient literature concerning the effects of SES and food environments on energy balance and diabetes, two conceptual relationships were investigated, neighborhood SES effect and food access effect on the outcome of diabetes.

These relationships are of key importance in this study because they have different influences on energy balance as seen by their effects on maintaining a healthy diet, and thus may directly or indirectly influence the outcome of diabetes. Based on these conceptual relationships, variables that were hypothesized to be related to the diabetes outcome were identified. Using the various databases described in the previous chapters,

[16] Cronbach's Alpha for Food Outlet Density was 0.640.

these variables were extracted and spatially linked to the addresses of the study subjects in order to tackle the following specific aims of the study:

Aim [A] to measure small area food environments around the study subject's residences

Aim [B] to characterize distribution of diabetes cases in Brooklyn

Aim [C] to analyze data for the presence of spatial clusters of diabetes in the study area

Aim [D] to measure the effect of food and socio-economic environments on a diabetes outcome

Before an explanatory analysis could be conducted, a spatial analysis methodology to measure neighborhood SES and spatial access to food outlets needed to be developed. The main goal was to spatially link individual-level health data, extracted from patient's individual Electronic Medical Record, to the neighborhood SES and to an index of food access. This step was critical to producing the reliable statistical estimates of the relative effect of SES and food environment around the subject's residences on the diabetes outcome later. To accomplish this, for each subject in the dataset, a dichotomous measure of the individual's status with respect to the outcome of interest (diabetes) and demographics were recorded. Several neighborhood-level SES environment variables were constructed from the US Census 2000 Summary Files 1 and 3 (SF1 and SF3). First, the SES data was aggregated up to the Census Tract of the subject's residence. These variables included poverty, percentage of black residents, markers of formal educational attainment, and population density per neighborhood of residence. Table 7 characterizes neighborhood socio-economic indicators abstracted from US Census 2000 and aggregated to CT's of residence of the subjects.

Socioeconomic Characteristics of Neighborhood of Residence by Diabetes Status						
Socioeconomic indicators at Census Tract level (SF1 and SF3, US Census 2000)	*Mean*		*Standard Deviation*		*t-test*	*p-value*
	Diabetes	*No Diabetes*	*Diabetes*	*No Diabetes*		
Percent of population ages 25+ with high school education	63.44	62.65	9.99	10.21	3.705	.000
Percent of population ages 25+ with less than high school diploma	33.72	33.06	9.29	9.40	3.327	.001
Percent of population ages 25+ with less than 9th grade	10.87	10.78	4.17	4.32	.966	.334
Percent of population below the Federal poverty line	37.93	37.11	13.64	13.55	2.841	.005
Percent of housing units with no private vehicle	62.61	61.70	15.21	15.51	2.822	.005
Percent of black residents	5.85	7.11	11.54	14.17	4.811	.000
Percent of female-headed households with children	19.94	19.42	6.92	7.19	3.505	.000
Total population density her square mile	63,822.00	63,548.00	28,937.87	29,554.91	.442	.658

Table 7. Socio-economic characteristics of CT of residence by diabetes status.

SES data were then joined with the subject-level data, access to the food stores data, and restaurants that were most proximate to each individual subject's residence. The densities of each type of outlet (e.g., fast-food outlets, fine-dining restaurants, and supermarkets) per neighborhood were also included.

Aim [A] Measurement of Small Area Food Environments Around the Study Subject's Residence

The in-outlet food environment was proxied by the detailed coding of all 9,739 licensed food outlets in Brooklyn into 10 categories (e.g., fast-food outlets, fruits and vegetables markets, fish markets, meat stores, etc.) [17], using the food outlet coding approach described in the previous chapter. Due to the concerns about the accuracy of the buffer methods raised by other researchers (see review in Chapter 3) an alternative, raster-based approach of measuring spatial access to food outlets, was developed and tested. This alternative approach makes it possible to calculate the walking times from the subject's home, anywhere in the study area, to the closest location for each of the 15 types of food outlets. Modeling spatial access to food outlets by walking involved the creation of a raster GIS model which required several steps. In the first step of this process, data points for each type of food outlet were coded (refer to Tables 2 and 3) and grouped into somewhat broader categories that were presumed to offer similar food choice options. Raster grids for each of the groupings were then generated in the second step. Finally, the resultant surfaces were used as inputs for the cost-weighted distance formula discussed later in this chapter.

The Creation of the Walkable Street Grids

In order to prepare the shapefile of walkable Brooklyn streets for processing several steps needed to be taken. To create the input friction surface, the shapefile of Brooklyn streets was extracted from the 2009 LION geodatabase.[18] To ensure that only streets open to pedestrians were used for walking access matrix generation, the following types of

[17] 5,313 food service establishments and 4,426 food stores were successfully geocoded.

[18] 2009 LION geodatabase was produced by the New York City Department City Planning.

features were removed before converting the vector-based file of streets to the raster grid[19]:

"NonPed" = 'V' OR "FeatureTyp" = '1' OR "FeatureTyp" = '2' OR "FeatureTyp" = '3' OR "FeatureTyp" = '4' OR "FeatureTyp" = '7' OR "FeatureTyp" = '8'

where:

- *"NonPed" = 'V'* -- Vehicle-only: primarily roadways, inaccessible to pedestrian usage
- "*FeatureTyp" = '1'* -- Railroad
- "*FeatureTyp" = '2'* -- Water Edge/Shoreline
- *"FeatureTyp" = '3'* -- Census Block Boundary
- *"FeatureTyp" = '4'* -- Other Non-Street Feature: Physically existing, addressable boundary, such as a street segment that has been closed off (inaccessible at both ends).
- *"FeatureTyp" = '7'* -- District Boundary: Physically non-existent boundary for a community district, a police precinct, or a fire company.
- *"FeatureTyp" = '8'* -- Physical Non-Street Boundary: Physically existing un-addressable boundary (such as a rock wall cemetery edge).

The resultant shapefile was later converted into a raster grid and merged with the raster grid representing building footprints in Brooklyn. The logic of creating this model was based on the assumption that streets and public spaces which are open to pedestrians, without buildings as barriers, would generally be walkable. For the purposes of generating a GIS model, a surface of walkable streets and public spaces without buildings

[19] This SQL expression was used in Select by Attributes box in ArcMap 9.3 to remove unwalkable segments of streets.

were assigned values representing the cost of travel (measured in minutes of walking to the closest outlet) from any location in Brooklyn. This process is described in more detail in the following section.

The Calculation of Pedestrian Walking Speed and Distances

The mean walking speed of subjects was calculated based on the results of past methodology studies (Morabia et al. 2010; Morabia et al. 2009; Levine and Norenzayan 1999). These studies concluded that the mean walking speed of an adult pedestrian in New York City was about 10 minutes per 2,640 feet or 3 miles per hour. As part of the Morabia et al. study, which was conducted in New York City in 2008 (at the same time period for which the data collection for the present study occurred), walking times were verified by participant's journals and by cross-referencing these times with the GPS time stamps from the GPS units worn by the study participants. Other recent studies have corroborated the findings of Morabia et al. (Agrawal et al. 2008; Cervero 2006; Calthorpe 1993; Levine and Norenzayan 1999) that typically traveling 0.5 mile by foot with groceries takes approximately 15 minutes in urban settings. The urban planners also generally agree that food shopping and other routine activities typically take place within a 0.5--mile radius of the residence. Modeling the walking behavior of the study subject's access to the nearest food outlets was of great interest to this study. To generate an accurate understanding of the walking speed of a shopper, the number of minutes required to travel across the street surface was computed. Thus, a walking speed of 3 miles per hour would call for "friction" values of:

-1 hour per 3 miles, or

-20 minutes per mile, or

-20 minutes per 5,280 feet, or

-0.0037878 minutes per linear foot

All streets open to pedestrians in Brooklyn were converted into raster grid with a 5x5-foot pixel and a value of 0.018939 minutes per cell (0.00379 minutes x 5 feet in each cell). All other surfaces such as buildings, other built-up areas and any additional areas that would be off-limits for the casual walker (such as cemeteries and construction sites) were set to "No Data" value in the GIS model. The resulting friction grid was loaded into the Spatial Analyst extension of ArcMap and a Cost Distance operation was run. The Cost Weighted Distance operation, constrained by a special "friction" raster, traverses the cells that lay on the shortest walkable path to the closest food outlets.

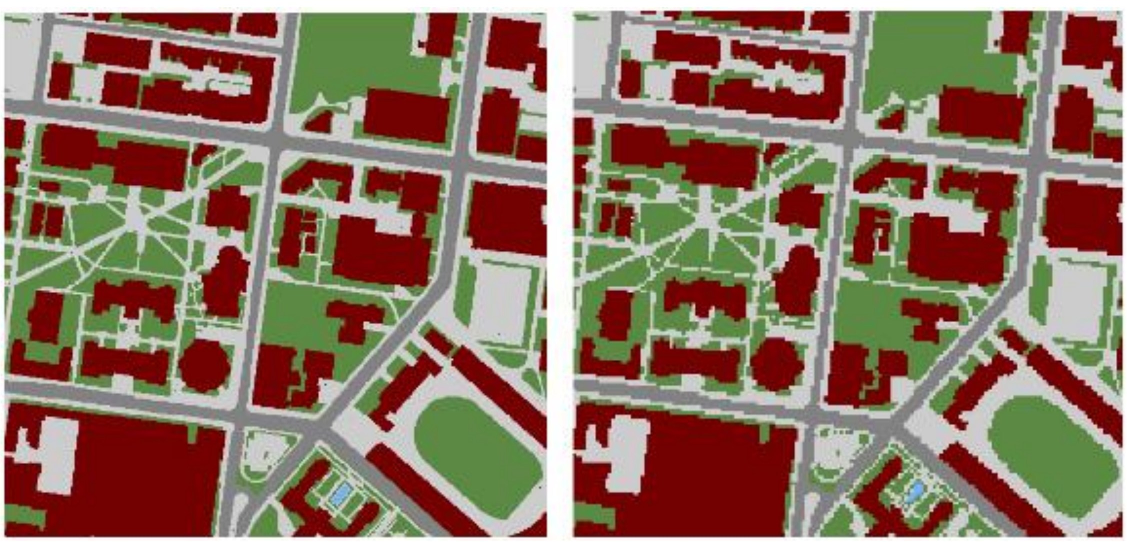

Figure 9. Raster image at 5 and 10 feet resolution per cell.

The operation produced a three-dimensional matrix or surface of cells where the "cost" of following the shortest (and cheapest in terms of minutes traveled) path was computed by adding the values of the cells in the "friction" raster that needed to be crossed (in this case the friction raster was the raster representing all walkable streets in Brooklyn). Thus, each point representing subject's home addresses could now be spatially linked with the cell value representing the cost of travel to each of the closest food shopping options (e.g., times to reach the closest supermarket, bodega, and fine food restaurant, etc.).

The Spatial Analyst extension in ArcMap Version 9.3 (Redlands, CA) and ET GeoWizards Version 10 (Pretoria, South Africa) were used to create the travel time surface. The details of that process are as follows:

1. The shapefile representing all walkable streets was converted into raster grid with 5x5--foot cell resolution and a cell value of 0.018939 and specified as the friction grid in the Cost Distance operation (CWD).
2. In order to convert the data from the shapefile to the raster grids one standard grid cell size needed to be established for all rasters. Given that the study area is approximately 10x10 miles or 52,800 x 52,800 feet, a cell size of 5x5 feet would result in a grid of 10,560 x 10,560 pixels. This pixel size was coarse enough to run the model reasonably fast on a desktop computer and fine enough to effectively distinguish between streets and sidewalks. Figure 9 shows two raster maps, one at a resolution of 5 feet and the other at a resolution of 10 feet (courtesy of Dr. Dana Tomlin, University of Pennsylvania).
3. The original point shapefiles, each representing all locations of the specific food outlets in Brooklyn (e.g., fast food vendors, fool service restaurants, bodegas, etc.

based on the classification developed in Tables 2 and 3) were used as destinations for the CWD operation.

4. The Cost Distance operation for each type of food outlet was then run. This step produced a series of distance-from-store raster surfaces for each type of food outlets coded.
5. The Raster Calculator was used to convert distance-from-store into an accessibility index for each type of food outlet based on the mean walking speed of 3 miles per hour. For instance, on a 0-100 minute scale, a subject's spatial accessibility to a supermarket would be highest (0 minutes in travel cost) if the subject's residence is within the same building as the supermarket but then could increase to 35 minutes of walking or greater, per se, if the subject lives in an area with no supermarket coverage. Importantly, this increase may be non-linear as the path winds its way along the cellular cost surface to the store. All walking paths avoided any non-walkable surfaces, such as large highways without pedestrian access, built up areas and closed alleyways, thus mimicking the pedestrian behavior as closely as possible.
6. The resulting accessibility index grids were then overlaid with the point shapefile the geocoded residences of subjects. ET GeoWizards software was used to snap the points representing the geocoded addresses of subjects to each of the accessibility index grids. The travel cost values (in minutes) of the cells closest to the residence location were extracted to subjects using the Extract Values to Points function in the Spatial Analyst extension.

7. This operation was then repeated for all the food types and several groupings of food outlets based on Tables 2 and 3. The distance between the point representing each subject's residence and the point representing the closest food outlet was expressed as walking time in minutes. Figure 10 shows the final map output of the CDW operation for supermarket access for the entire study area denoted in the enlarged area in the lower right corner. This insert shows the detailed map of the enlarged area with 2 supermarkets, housing stock footprints, parks and other unwalkable surfaces (shown in black). The cost-weighted distance raster (yellow representing closer locations and orange representing farther locations) was used to determine the time it would take for each case and control in the study area to reach the closest supermarket by foot. Using the Extract to Points tool in the Spatial Analyst, the points representing the geocoded addresses of subjects, masked for confidentiality, were given the closest raster cell value representing walking time in minutes to the closet supermarket. From this map it is obvious that many Brooklyn residents live outside the .5 mile desirable walking radius from the closest supermarket. The same process was repeated for all the other food outlets.

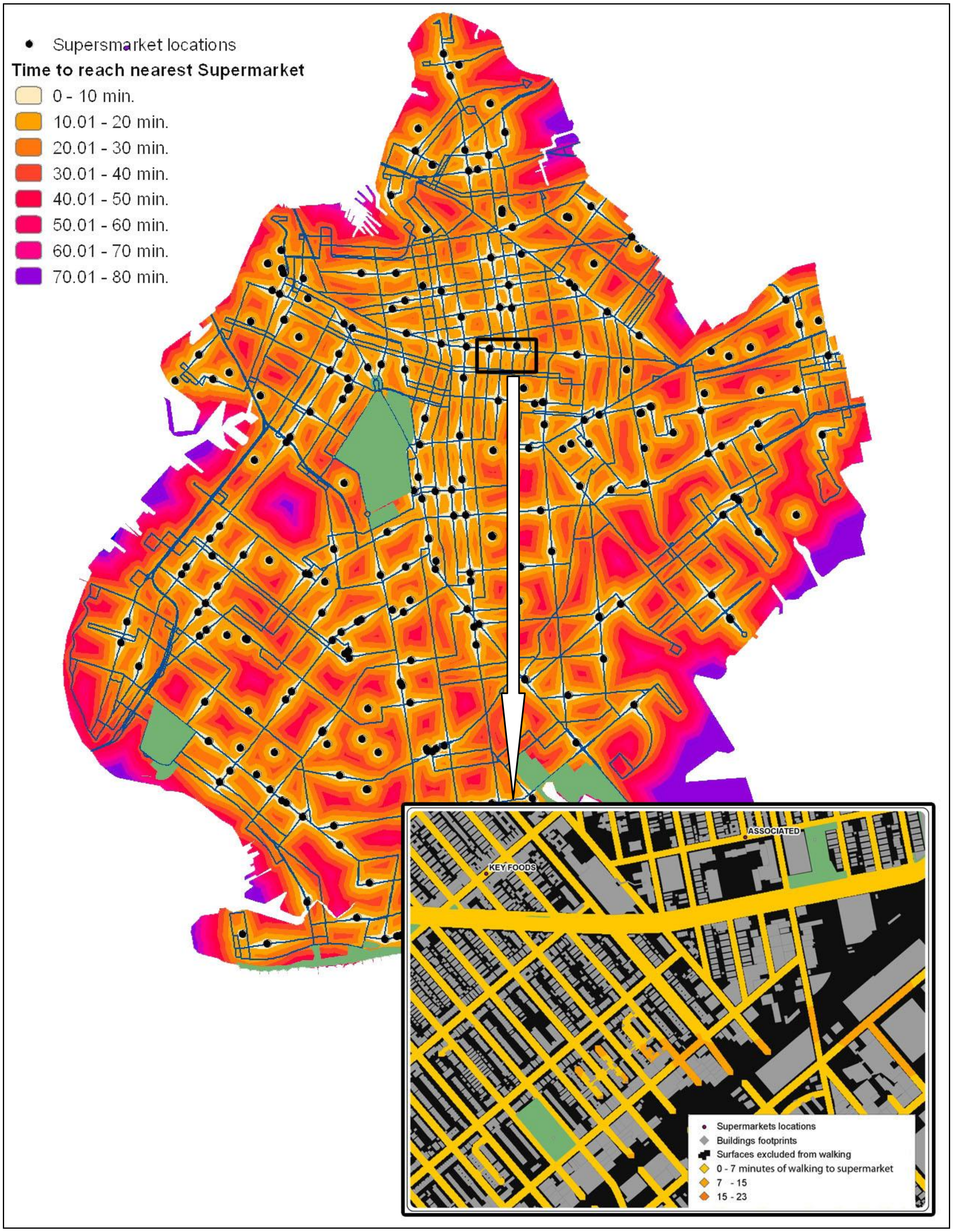

Figure 10. CWD calculation for supermarket access

The resulting output from the GIS model was then used in multilevel logistic regression equations, discussed in the Statistical Analysis section of this chapter. For the purpose of the regression, each subject's residence cell was treated as a single statistical observation dichotomized to the outcome (having or not having diabetes). The process of constructing a specific model using these data is described in the next sections.

Characterization of the Food and SES Environment

The following tables provide characteristics of the local food environment by type of outlet and by the density of outlets, dichotomized by the presence of the outcome of diabetes by subject. These initial descriptives proved to be helpful in characterizing the study population in respect to the outcome of interest. More specifically, Table 8 represents some of the characteristics of the local food environment of subjects by diabetes status while Table 9 represents individual demographic characteristics including age, race/ethnicity, and the gender of the sample by diabetes status. Residence in public housing is also described. All variables are dichotomized by diabetes status. Because socio-economic data is not collected during hospital visits, aggregated socio-economic and demographic data from the US Census Bureau, Census 2000, were employed to model the social and demographic environments in the neighborhoods of residence of the cases and controls. It is difficult to tell if there are any significant differences between the subjects and controls in accessibility to different food outlets in Brooklyn based only on the results of descriptive statistics alone (t-test for continuous variables and Chi-square test for categorical variables). Therefore, a more sophisticated approach was required.

Characteristics of local food environment of subjects by diabetes status (study period 05/01/08 to 10/30/08) n=10,720						
Food environment variables	*Mean walking times in minutes*		*Std. Deviation*		*Tests of Independence*	
	Diabetes (n=3,202)	*No diabetes (n=7,518)*	*Diabetes*	*No diabetes*	*t-test*	*p-value*
Mean Walking Distances to Food Outlets						
Walking time to supermarkets	5.82	5.87	3.32	3.39	.818	.414
Walking time to independent grocery stores	2.11	2.17	1.68	1.80	1.554	.120
Walking time to fruits and vegetables markets	8.56	8.54	6.70	6.63	.132	.895
Walking time to convenience stores	3.53	3.56	2.14	2.18	.659	.510
Walking time to convenience stores and bakeries	2.02	2.03	1.43	1.46	.419	.675
Walking time to meat and fish markets	5.90	5.90	3.67	3.72	.040	.968
Walking time to bakeries and confectioners	12.56	12.39	6.78	6.56	1.201	.230
Walking time to cafeterias and full-service	5.72	5.62	3.55	3.49	1.376	.169
Walking time to fast-food restaurants	2.21	2.20	1.52	1.53	.274	.784
Walking time to bars and pubs	14.66	14.29	9.44	9.32	1.86	.063
Total count of all food outlets per CT	13.39	13.11	9.57	9.36	1.426	.154
Mean Densities by Type of Outlets per Mile² (in CT of residence)						
Total Density of all food outlets combined	187.62	185.08	129.45	128.26	1.413	.158
Density of supermarkets	5.68	5.62	8.68	8.55	.352	.725
Density of independent grocery stores	54.97	54.31	40.99	40.82	-.767	.443
Density of fruits and vegetables markets	4.65	4.58	8.73	8.66	.369	.712
Density of convenience stores and delis	15.76	15.85	16.79	17.32	.230	.818
Density bakeries, confectionery and nut stores	0.85	1.02	3.61	4.01	2.075	.038
Density of cafeterias and full-service restaurants	7.87	8.61	13.35	15.41	2.384	.017
Density of all fast-food restaurants	83.72	81.16	63.26	62.39	1.937	.053

Table 8. Characteristics of the local food environment of subjects by diabetes status.

Demographic and housing characteristics of subjects by diabetes status				
Subject Variables	*Status*		*T-tests of Independence*	
	Diabetes (n=3,202)	*No diabetes (n=7,518)*	*χ2*	*p-value*
Age in years, SD	61.74	53.40	33.744	<.001
Gender n,% of sample size.				
Female n=6239, 58.2%	2014	4225	41.433	<.001
Male n=4481, 41.8%	1188	3293		<.001
Race/ethnicity n,%				
Black, 9,756 (91.0%)	2,968	6,788	3.755	<.001
Hispanic 502 (4.7%)	145	357		<.001
White 148 (1.4%)	19	129		<.001
Other 314 (2.9%)	70	244		<.001
Type of Housing by Subject				
Residence in Private Housing	2,947	7,043	9.583	0.002
Residence in Public Housing	255	475		<.001

Table 9. Demographic and Housing Characteristics of Subjects by diabetes status

Measures of Food Outlet Density

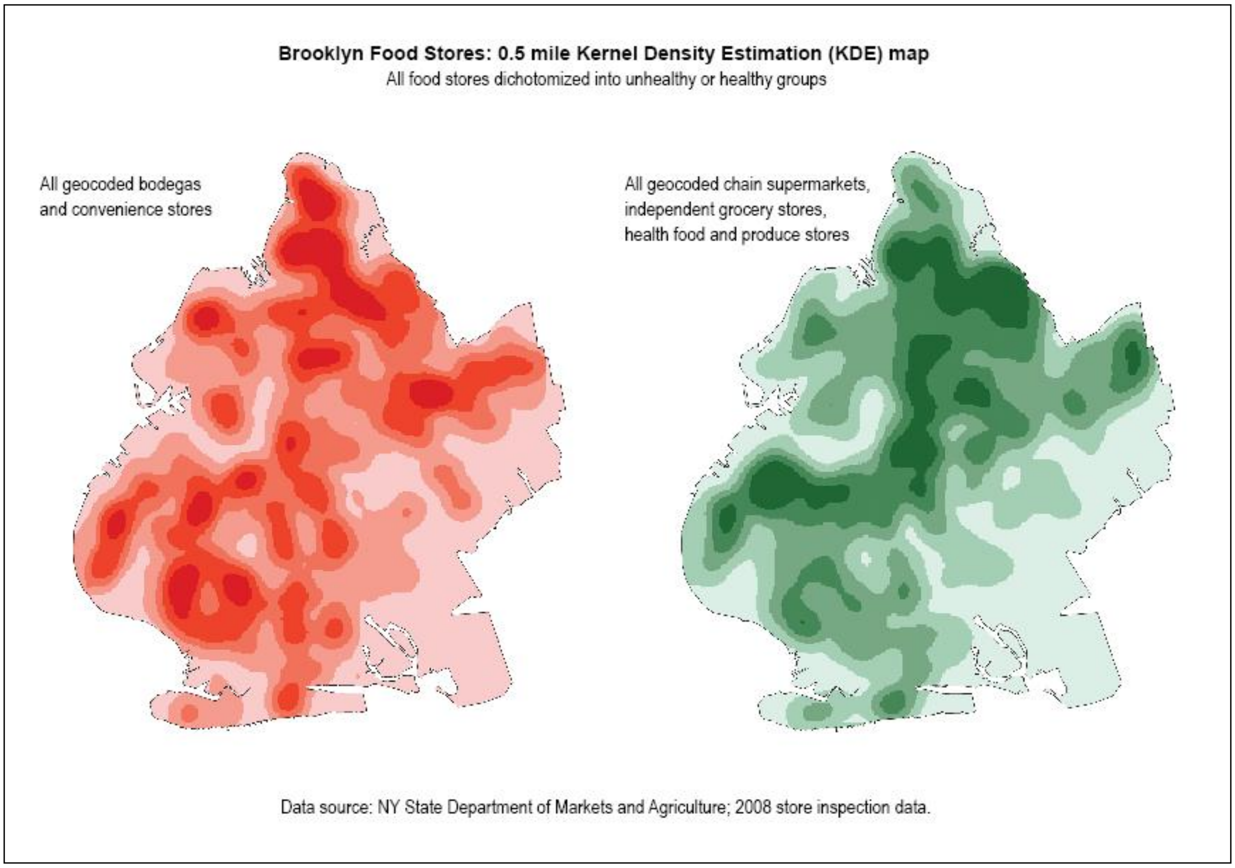

Figure 11. The maps show kernel density estimates of food outlet densities by type using a .5 mile bandwidth.

The two maps in Figure 11 show the stark differences in total food access among various neighborhoods. Overall, the results of the initial mapping of food environments using multiple kernel density estimates (KDEs) showed that metabolism-healthy and unhealthy outlets were unevenly distributed in Brooklyn. KDE is a cartographic technique that calculates the spatial densities of variables of interest within a specified search radius or bandwidth. A bandwidth of 0.5 mile was used in both Figure 12 maps. This was based on the estimates that typical food shopping behaviors usually take place within a 0.5 mile radius from an individual's residence (Agrawal at al. 2008; Calthorpe 1993; Cervero 2006). The mapped outlets were dichotomized by metabolism-healthy and unhealthy food outlets. This distinction was made based on the availability of fresh produce, meat, and fish verses fast foods. Based on this mapping, it appears that areas with low healthy food store density are dominated by the unhealthy outlets, as exhibited in the map of the left. Importantly, there are parts of eastern Brooklyn that have virtually no food outlets, of any type, within a 0.5 mile radius from the study subject's residence.

In addition to the individual subject's walking access to different types of food outlets described above, food outlet density was also hypothesized to be an important factor in dietary choices. In order to measure food density and include it in the analysis, food density indices were computed for all 485 census tracts where study subjects lived at the time of the data acquisition in 2008. In 2008 Brooklyn had 5,562 food outlets, of all types, with a mean outlet density of 11.5 (SD 9.65) outlets per census tract. Interestingly, densities for most types of food outlets covaried with the densities of other retail businesses across the study area. The highest density of food outlets was in the census tracts containing high densities of other retail businesses with at least one major urban thoroughfare. The highest outlet density, 901 food outlets per square mile, was found in a busy, racially and ethnically diverse and densely populated census tract in a downtown area in Brooklyn, where the mean number of outlets was 165.7 (SD 144.05) per CT. The least food-dense CT was located in the New Lots area of Brooklyn, where the population is predominately African-American. The mean food density in this area was only 5 outlets, of any type, per square mile. This highly uneven spatial distribution of food outlets highlighted the stark differences in the spatial access to food across the neighborhoods. Figure 12 graphs the distributions of all coded types of outlets per CT. Approximately 50 CT's had a very high food outlet density, 900 outlets per square mile, providing a very high number of food choices, while the majority of CT's had much fewer food options.

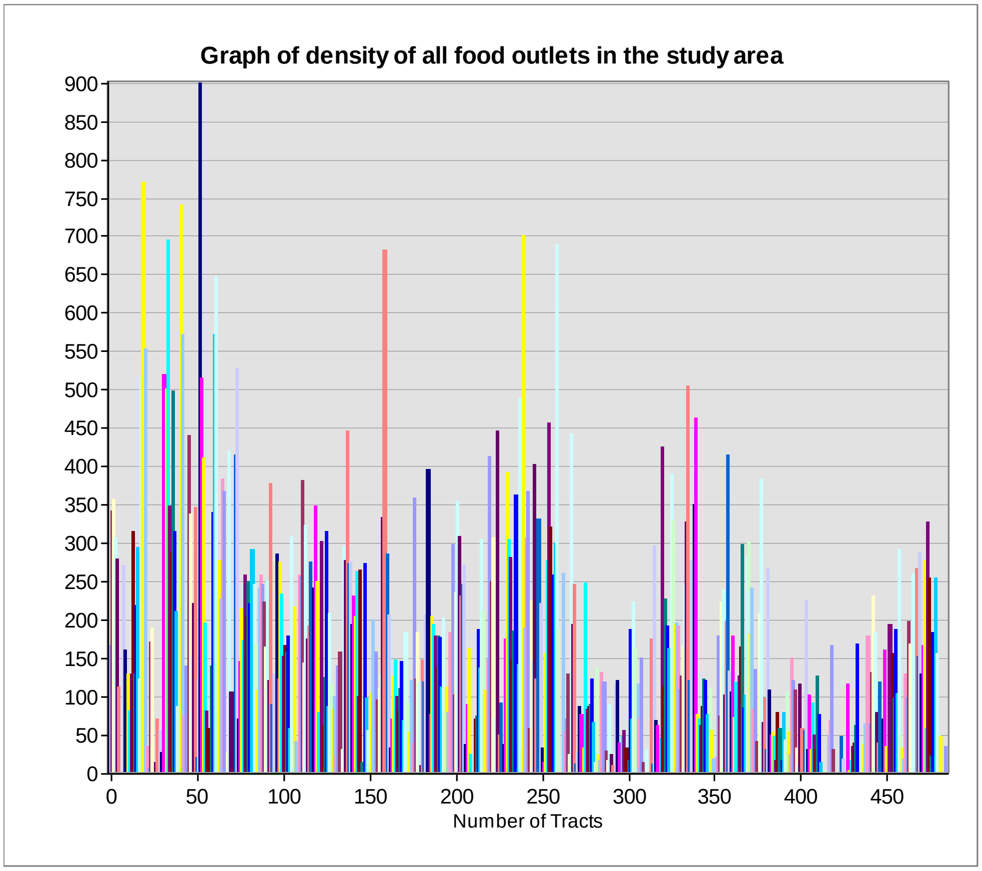

Figure 12. Graph of food density (all types of licensed outlets) per CT

Figures 13 and 14 below further illustrate this critical issue. Figure 13 provides a visual example of an urban area with high food density tracts. The area highlighted in bright blue denotes boundaries of just one CT in a relatively high-income area of Brooklyn.

Residents of this CT clearly have ample choices of healthy food options and easy spatial access to virtually any type of food outlets coded.

In contract, Figure 14 provides an example of a near “food desert”, located in an impoverished CT in the New Lots area of Brooklyn. Importantly, this CT has roughly the same population as the CT shown in Figure 13. It should be noted that not only is there a low food density but also a lack in the diversity of food choices in this CT. The local resident’s food shopping choices here are severely limited to just a few corner delis and bodegas, selling mostly soft drinks, high-calorie snacks and other fast food items. No dedicated produce stand was found by the author inside of any of the outlets mapped in Figure 14 during a field visit on 10/12/2010.

It is plausible to think that in commercially desirable areas such as the one in illustrated in Figure 13, the spatial density of the food outlets creates conditions in which the local stores and restaurants are forced to compete for customers and for the limited commercial space available. This economic pressure is likely to be responsible for the variety of food choices available in CT’s similar to the one shown in Figure 14. Inversely, in CT’s with similar to the CT shown in Figure 14, the economics of the food retail industry may create conditions which make it more profitable to sell easily-marked up items such as soft drinks and snacks, as opposed to items such as fruits and vegetables.

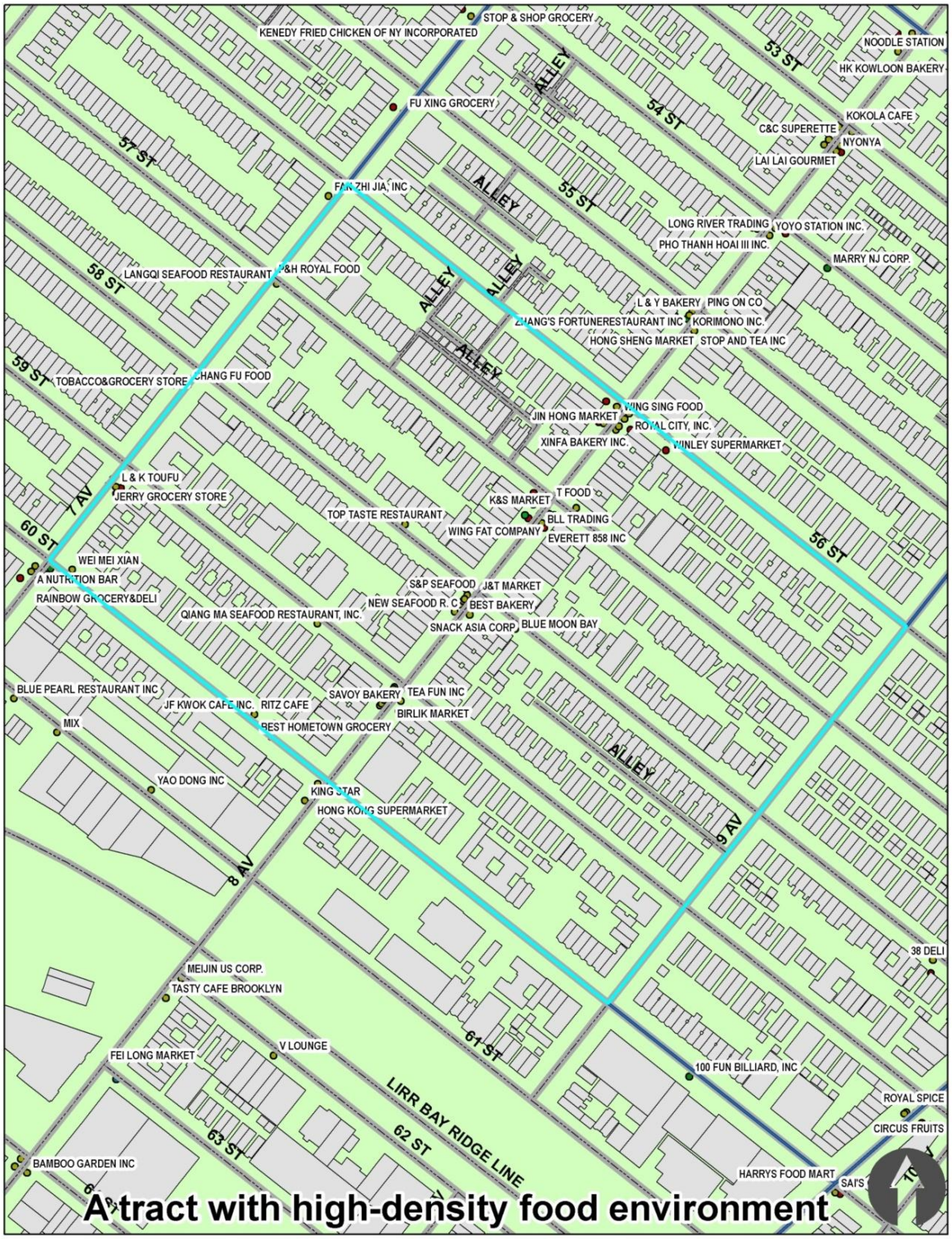

Figure 13. High Food Density Tract.

Figure 14. Very low food density tract with roughly the same population.

Aims [B]and [C]: To Characterize the Spatial Distribution of Diabetes Clusters in Brooklyn

The initial mapping of the spatial distribution of diabetes cases from the sample showed an increased prevalence of diabetes in neighborhoods located in Central and Eastern Brooklyn. Based on the NYC Community Health Survey data, all of these areas have the highest prevalence of diabetes in Brooklyn and some of the highest in New York City (Figure 1). In order to further explore the spatial pattern of diabetes, it was important to assess whether the diabetes cases are randomly distributed across the study area or if they exhibit a non-random spatial pattern. Traditional cartographic methods of representing the cases as points or by aggregating the points to areas to produce a choropleth map have severe limitations in respect to visualizing disease clusters. For what may appear to be a visible diabetes cluster in a particular part of the map, can often be explained simply by a clustering of the total population in that area due to housing stock density. Therefore, a search for an alternative process to measure clustering, responsive to these unique constraints of spatial data, was performed.

Kulldorff's Spatial Scan Statistic

As Kulldorff showed in his 1995 paper, the detection of the spatial clustering of disease must avoid ad hoc measures such as mapping points or prevalence of disease to detect clusters within the neighborhoods studied. Therefore, a Kulldorff local spatial scan statistic was used to search for spatial clusters of diabetes. Unlike a simple mapping of cases and other arbitrary methods of cluster visualization, the local spatial scan detects clusters of disease cases after adjusting for spatial variations in the density of the background population using the Poisson model or, as in the case of this study, using case—control data in the Bernoulli model.

The local spatial scan produces a likelihood ratio statistic that detects the location of the possible disease clusters in a population with inhomogeneous spatial density, while simultaneously uses methods of inference to test for significance. This approach was already successfully used in the geographic analysis of diabetes clusters in an urban area of Manitoba, Canada to detect the presence of clusters (Jemal et al. 2000). After reviewing the salient literature and available software on detecting the spatial clustering of disease (Jemal et al. 2000; Kulldorff, Tango, and Park 2003; Kulldorff et al. 2006), the SatScan 9.1, a free software developed by Martin Kulldorff at the Harvard Medical School, was used.

In studies where the incidence of a particular outcome is known, the local cluster search produces output that measures clustering in units of risk (likelihood) ratios (RR). RRs are produced by calculating the estimated risk within the cluster and then dividing it by the estimated risk outside the cluster. It is calculated as the observed number of cases divided by the expected number of cases within the cluster divided by the observed cases divided by the expected outside the cluster. Since this analysis is conditioned on the total number of cases observed, E[C]=C):

$$RR = \frac{c / E[c]}{(C - c)/(E[C] - E[c])} = \frac{c / E[c]}{(C - c)/(C - E[c])}$$

The Bernoulli model, which is the most appropriate model for this case/control study and other studies with 0/1 event data was run in SatScan 9.1 software to determine if there are spatial clusters of diabetes within the study sample. The likelihood ratio tests were performed, and for each location and size of the scanning kernel with alternative hypothesis that there is an elevated risk within the kernel as compared to outside. When the point data representing residences of cases and controls was analyzed for statistically significant local clusters, a number of clusters with high case/control and prevalent cases were detected (see summary below).

Example of primary cluster statistics:

```
Semiminor axis........: 34397.53
Semimajor axis........: 34397.53
Angle (degrees).......: 0
Shape.................: 1.00
Population............: 3201
Number of cases.......: 3201
Expected cases........: 97.92
Observed / expected...: 32.69
Relative risk.........: 101476.00
Log likelihood ratio..: 14305.698143
Test statistic........: 14305.698143
P-value...............: < 0.00001
```

The RR map in Figure 15 shows the results of the five levels of statistically significant diabetes clusters in the study area aggregated to the CT of residence. This data provides a much higher spatial resolution of the location of hot spots of diabetes as compared to any other NYC studies, such as NYC HANES, which use a much coarser level of aggregation. The RR clusters for each CT produced by SatScan show that there is a statistically significant cluster of diabetes cases in central Brooklyn, as seen in the middle of the map.

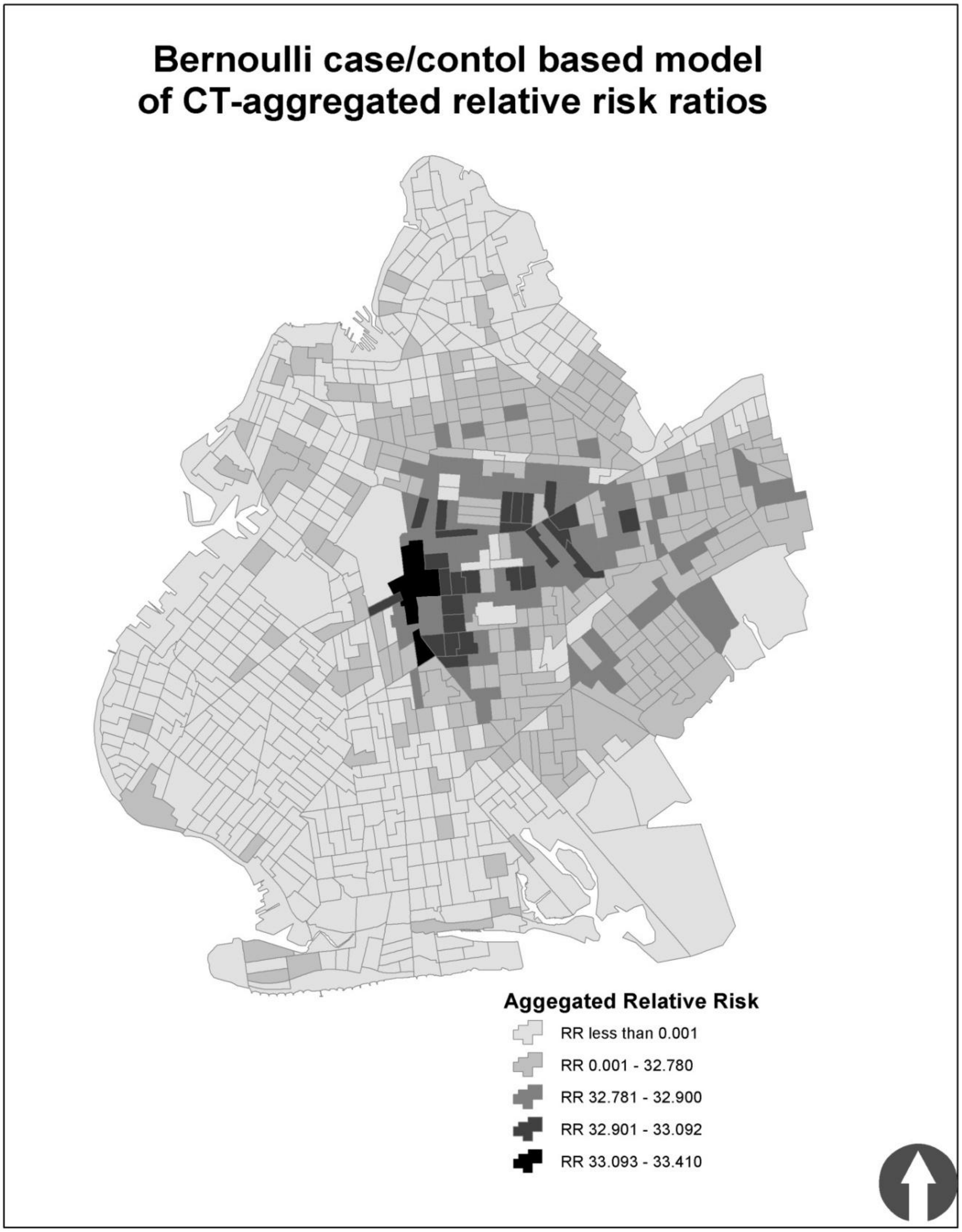

Figure 15. Results of the five levels of statistically significant diabetes clusters aggregated to CT of residence

Aim [D]: To Measure the Effect of the Food and Socio-Economic Environments on a Diabetes Outcome

Multilevel Analysis

After aims A, B and C were accomplished, a case-control study design was employed. Several multi-level models were fitted with diagnosed diabetes as the dependant variable, and the independent variables of neighborhood SES and the time it takes to access various food outlets (scored on the relative variety of food options offered inside the outlet, including the in-outlet availability of fresh fruits and vegetables) from home.

Based on the insights from the salient literature and from the exploratory analysis using SatScan spatial scan statistic, several multilevel logistic regression models were constructed. SAS version 9.3 (SAS Institute, Gary, NC) and R version 2.3 were used in the analysis. The goal was to create composite indices for food and SES measures. Similar to the SES data, there was also a strong correlation in access to food outlets, which led to the performance of a factor analysis on food access and food outlet density variables. Using the outcome of diabetes as the dependant variable, and the output of factor analysis on SES, food access and density as independent variables, a set of logistic regression models were fitted.

The initial simple logistic regression in Model 1, shown in Table 9 included only the individual level variables of age (divided into four quintiles) gender and walking distances to food outlets. Deviance residuals were Min =1Q; Median=3Q, Max:-1.3824,

-0.8459, -0.5926, 1.1038, and 2.0207, consecutively with null deviance of 12,060 on 9788 degrees of freedom and residual deviance of 11,040 on 9782 degrees of freedom (1 observation deleted due to missingness). The Akaike information criterion (AIC) was 10945.5, and the number of Fisher scoring iterations was 4. Female gender and older age were both strongly associated with a diabetes outcome in the subjects, while the latter variable of food access by walking shows no significant association with this outcome. The following four age groups were constructed for the Model: Age 50-59 years; 60-69 years; 70- 79 years and; 80 years and older. In this initial Model none of the neighborhood-level SES or food predictors were included. As expected, the results showed that old age and gender are statistically significant predictors of diabetes. In this Model, dispersion parameter for binomial family was taken to be 1.

	Model 1	OR (95% CI)	P-value	Significance
Level 1	Female gender	1.26 (1.14-1.38)	<0.0001	***
	Age Group 50-59 years	2.36 (2.08-2.67)	<0.0001	***
	Age Group 60-69 years	4.94 (4.33-5.64)	<0.0001	***
	Age Group 70-79 years	6.54 (5.57-7.68)	<0.0001	***
	Age Group 80+ years	8.63 (7.00-10.67)	<0.0001	***
Goodness-of-Fit				
AIC		10945.5		
BIC		10995.8		
Log Likelihood		-5465.8		
Deviance		10931.5		
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Table 10. Model 1.

The variance Inflation Factor (VIF), illustrated in Table 11 below, was very close to 1. Cronbach's Alpha was relatively high at 0.7272795. Goodness-of-Fit measures are a way to describe how well the Model fits this set of observations. While there are many different ways to measure the goodness-of-fit, the Model used in R relies on AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), Log-likelihood and Deviance. However, the detailed interpretation of the goodness of fit measures is well outside the scope of this study.

Variance Inflation Factor (VIF)			
	GVIF	Df	GVIF^(1/(2*Df))
Sex	1.005091938	1	1.002542736
Age Group	1.004692792	4	1.000585398
Food Environment	1.001578611	1	1.000788994

Table 11. Variance Inflation Factors

The subsequent Models were fitted with the dependant variable of an outcome of diabetes and individual level variables (level 1) as well as neighborhood SES measure (level 2). Model 2 (Table 12) included individual-level variables used in Model 1, food outlet access predictor and neighborhood-level SES predictor. Model 2 shows that neighborhood socio-economic deprivation predicts diabetes when the individual-level variables are held constant. Model 2 included the same individual-level variables as Model 1 but also included CT-level socio-economic deprivation. No food access measure was included in either Model.

	Model 2	OR (95% CI)	P-value	
Level 1	Female Gender	1.26 (1.15-1.38)	<0.0001	***
	Age Group 50-59 years	2.37 (2.09-2.68)	<0.0001	***
	Age Group 60-69 years	4.97 (4.36-5.68)	<0.0001	***
	Age Group 70-79 years	6.59 (5.61-7.73)	<0.0001	***
	Age Group 80+ years	8.70 (7.04-10.74)	<0.0001	***
Level 2	Tract SES	1.08 (1.03-1.13)	0.0016	
Goodness-of-Fit				
AIC		10937.6		
BIC		10995.1		
Log Likelihood		-5460.8		
Deviance		10921.6		
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 12. Model 2; The Effect of individual-level variables on the outcome of diabetes with added neighborhood socio-economic deprivation measures.

Model 3 included the same individual-level variables as Models 1 and 2 but also included a CT-level food access measure. While a significant change in the Model was observed when an SES variable was added in Model 2 the neighborhood SES, showing that neighborhood SES is a statistically significant predictor of diabetes, no such change occurred when the neighborhood food access measure was added (Model 3).

	Model 3	OR (95% CI)	P-value	
Level 1	Female Gender	1.26 (1.14-1.38)	<0.0001	***
	Age Group 50-59 years	2.35 (2.08-2.66)	<0.0001	***
	Age Group 60-69 years	4.94 (4.33-5.64)	<0.0001	***
	Age Group 70-79 years	6.54 (5.57-7.68)	<0.0001	***
	Age Group 80+ years	8.65 (7.00-10.68)	<0.0001	***
Level 2	Tract Food Density	1.04 (0.99-1.09)	0.1167	
Goodness-of-Fit				
AIC		10945.1		
BIC		11002.6		
Log Likelihood		-5464.5		
Deviance		10929.1		
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 13. Model 3; The Effect of individual-level variables and food outlet density variables on the outcome of diabetes.

Model 4 (Table 14) showed no statistically significant effect of the localized food environment on the outcome of diabetes. However, the SES index in the neighborhood of residence in Model 4 still predicts higher odds of a diabetes outcome. Specifically, for every unit change in the Neighborhood SES, the subjects were 7% more likely to have diabetes (95% CI 1.03-1.12 at p-value 0.0024).

	Model 4	OR (95% CI)	P-value	Significance
Level 1	Female Gender	1.26 (1.15-1.38)	<0.0001	***
	Age Group 50-59 years	2.36 (2.09-2.68)	<0.0001	***
	Age Group 60-69 years	4.97 (4.35-5.68)	<0.0001	***
	Age Group 70-79 years	6.59 (5.61-7.73)	<0.0001	***
	Age Group 80+ years	8.7 (7.04-10.75)	<0.0001	***
Level 2	Tract SES	1.07 (1.03-1.12)	0.0026	**
	Tract Food Density	1.03 (0.98-1.09)	0.2097	
Goodness-of-Fit				
AIC		10938.0		
BIC		11002.7		
Log Likelihood		-5460.0		
Deviance		10920.0		

Table 14. Model 4, Individual level and neighborhood level variables

In summary, all Models included the dependant variable of diabetes, subject's demographics, as well as specific combinations of variables measuring the local food environment and neighborhood SES. The results indicate that low neighborhood SES, female gender and older age have a strong statistical association with diabetes. These findings are consistent with the general consensus in the literature that low SES, female gender and age are risk factors for diabetes. There was no evidence of a statistically significant association between food density and the outcome of diabetes.

Importantly, the findings show that the OR for individuals who live in the neighborhoods with lower SES is slightly higher for having diabetes. Based on the results of the multilevel logistic regression analysis, the effect of the aggregate index of social and economic deprivation in the neighborhood of residence was shown to have moderate but statistically significant associations with a diabetes outcome.

Chapter 5: Discussion, Limitations and Lessons Learned

Introduction

Results of the multilevel logistic regression discussed in Chapter 4 showed that a composite marker of socio-economic deprivation of the neighborhood of residence is statistically associated with the probability of having diabetes. This chapter focuses on the discussion of the results as they apply to diabetes and to other metabolic illnesses and on their meaning and implications in the context of the overarching goal of improving urban health. I also discuss the limitations of the data on which the analysis was based and of the explanatory models used in the analysis. The chapter concludes with the findings' policy implications for improving the prevention and control of diabetes in urban context. Future directions of this research are also discussed.

This research has its theoretical underpinnings in the human ecosystems paradigm, which suggests an ecologic relationship between the wellbeing of individuals and the totality of the human-constructed and natural environment in which we work and live (Grove and Burch 1997; Burch, Cheek, and Taylor 1972; McGinnis and Foege 2004; McGinnis and Foege 2000). So far, in accessing the impact of the physical (human-constructed and natural), social, and economic environment on the energy balance in the body, the majority of the work has been focused almost exclusively on the effects of various suspected offenders within the neighborhood on obesity. In fact, a relative dearth of literature addressing quantitative and spatial aspects on the impact of the neighborhood environment on diabetes served as an impetus for this study.

As a result, this study specifically centered on studying how a set of spatially-referenced explanatory variables within the neighborhood affect risk of diabetes, and therefore the overall thrust of this study differs from prior studies in energy balance-related problems. Most of the previous work on effects of socio-economic and food access on energy balance focused on obesity. This somewhat limited focus resulted in a relative dearth of peer-reviewed publications interrogating and attempting to explain clusters of diabetes in the context of SES and food access. The resulting gap in knowledge of the effects of context of neighborhood context (SES and food access) on clusters of diabetes served as an impetus for the research presented here.

Specifically, the current study attempted to tease out the magnitude of the effects of several spatially referenced SES and food access measures, and measure the influence these variables exert upon the probability of having diabetes. The results showed that individuals who live in neighborhoods with lower SES have higher odds of having diabetes and that older persons and persons of female gender are at greater risk of diabetes. These findings are interesting for several reasons. Firstly, they corroborate the previous research findings the low SES of the neighborhood has an effect on a higher probability of its residents having diabetes. Secondly, they show that the suspected linkage between poor spatial access to healthy food outlets and diabetes cannot be confirmed by this cross-sectional study. Thirdly, they underscore the importance of gender in the study of diabetes. In the following sections I discuss these findings in the context of the previous findings. I discuss the relative strengths and weaknesses of the key sources of data used in this study, and address the limitations of the current work and its future directions.

A Note on the Analysis Methodology

It is important to emphasize that to measure the neighborhood SES for this study a composite SES deprivation score was created using factor analysis discussed in Chapter 4. The method used CT-level SES variables, as markers of social and economic stability and cohesion computed for each subject's CT of residence. This resultant composite score yielded a measure of socio-economic deprivation in each of the CT boundary-delimited neighborhoods, with higher scores indicating higher levels of socio-economic deprivation in the CT. This methodology is not new and was applied in studies that examined social and environmental constructs that involve multi-level (hierarchical) structures, which typically preclude an application of spatially weighted regression (Rundle et al. 2007; Rundle et al. 2009). This approach has been effectively applied in other studies of neighborhood health where one composite measure is often needed to fit multilevel regression models (Rundle et al. 2009; Auchincloss et al. 2008; Daniel et al. 2008). The methods used in this study required the derivation of 5 CT-level SES indicators as well as measures of spatial access to food outlets to as many as 7 types of different food outlet types for each of the subjects. As a result, the process required for construction of the dataset to be used in this 2-tier HLM was lengthy. However, this exercise yielded rewarding results in that it provided me with valuable knowledge and experience, easily transferrable to other geographic contexts and research questions which consider the effect of place on individual wellbeing, discussed in this chapter.

Diabetogenic Effects and Neighborhood Socio-economic Deprivation

One of the conclusions of this study is that neighborhoods with high levels of socio-economic deprivation have diabetogenic effect on its residents regardless of type and density of available food outlets. Such effect may be due to a number of factors that need to be explored. Individuals living in neighborhoods with low SES may be exposed to more stress that these living in higher SES neighborhoods. Since high stress environment has been shown to be a factor in the development of diabetes and of other metabolic disorders (Heraclides et al. 2011), this relationship should be explored in future studies. Previous studies corroborated that while various food outlets may be present in inner city neighborhoods; these outlets typically have low availability of healthy foods and the quality of foods and quality of the products that are available is often poor. Because of the lack of more precise in-outlet food environment data, this study assumed that all outlets which belong to a particular category (e.g., fast food restaurants or fruits and vegetables markets) have the same effect on metabolism, regardless of the SES of the neighborhood where they are located. As stated in previous chapters, it would be an oversimplification to presume that access to healthy food is defined only by spatial access to outlets. This is because of the differences in quality and availability of healthy food options inside stores of the same general type. For example, Morland showed that in addition to the effect of spatial assess to food outlets, the in-outlet[20] food environment

[20] Survey-based empirical data on which food options are available in local stores and food establishments.

should be assessed carefully (Morland and Evenson 2009). These researchers found significant differences in the quality and variety of fruits and vegetables in wealthier, predominately white areas in New York City as compared with poorer, minority areas. Even when comparing two supermarkets owned by the same corporation, one in a wealthier and predominately white area and another in a poor and predominately black area of NYC, these researchers found that the latter supermarket had lower quality and availability of foods considered diabetes-healthy (Morland and Filomena 2007). Specifically, they reported that choice, quality, and freshness of fruits and vegetables in poor CTs with large percentages of black residents was consistently lower while the prices were generally higher than in CTs with predominantly white population. The relationship between neighborhood SES and in-outlet food quality, choice and possibly prices inside stores and restaurants may have an effect on diets and on the outcome of interest (diabetes). The effect of the in-outlet food environment may be even more pronounced on the "captive" shoppers. The analysis presented here showed that these subjects living in the remote neighborhoods located away from the downtown shopping areas have lower densities of any type of food outlets near their homes. This means that these individuals need to travel longer to reach any food outlet. Importantly, food environment in these low-SES neighborhoods is characterized by a lower density of food outlets of any type. It consists mostly of small, independently owned delis, bodegas, and convenience stores, often co-located with gas stations, dominating the foodscape. These outlets offer mostly fast food type options, thus potentially limiting the food rations of the local residents to these diabetes-unhealthy options.

Another issue that should be examined in the future research is how access to public and private transportation affects food environment of the residents. While NYC has a robust public transportation system, low SES neighborhoods are sometimes located in outlying urban areas with low car ownership rates. Such is the case of the neighborhoods of eastern and northern Brooklyn, included in this study. These areas have both a significantly lower density of public transportation routes and lower frequency of bus and subway service. This existing urban condition creates a lose-lose situation in many American cities, especially for the residents with no access to private vehicles. They need to travel further away from home, either by walking or by using the infrequent public transit service to reach these more distant outlets providing healthy food options.

While this study did not collect data on individual diets, it is reasonable to assume that due to the spatial inequalities[21] in access to healthy food experienced on a daily basis by the residents of these areas, many individuals have little choice but to eat unhealthy foods, easily available in the outlets near their homes. The nascent research on neighborhood health needs to examine the totality of the existing transportation and food environment factors that may act as deterrents to achieving and maintaining healthy diets and lifestyles. For example, individuals often have to carry the purchased food back to their residences, healthier food options typically have a much shorter shelf life, are less energy-dense, and cost more per food calorie. They are also heavier and more bulky to carry. These factors may provide disincentives for buying diabetes-healthy foods for

[21] As compared with higher aggregate SES CTs had higher public transportation stops densities and more options to purchase healthy foods from the local outlets.

residents of low SES neighborhoods, and instead encourage them to buy highly refined and processed foods. In contrast, metabolism-healthy foods, such as fish, vegetables and raw grains, are typically less energy dense than fast food options and often cost more per pound (Patterson 2010).

Gender and Diabetes

The female subjects in this study had a 26 % higher probability of having diabetes than the males (CI 1.14-1.38, p=0.0001). This finding underscores the importance of examining the potential modifying effect of gender on diabetes. There is no general consensus on the role of gender in diabetes, but it is clear that the effect of gender on the outcome of diabetes changes with age, BMI and race/ethnicity(Heraclides et al. 2011). For example, the overall gender ratio is roughly equal in children diagnosed under the age of 15 but while European male adults show an excess of diabetes, middle-aged populations of non-European origin characteristically show a female excess of diabetes (Gale and Gillespie 2001). Some research on SES factors in diabetes also showed that women that are exposed to higher stress levels are at greater risk of developing diabetes than men who are exposed to higher stress (Heraclides et al. 2011). It may be the case in this study because the vast majority of the subjects in this sample lived in low-SES neighborhoods, they may have been exposed to a higher level of chronic stress, which has-gender-specific effect on the risk of developing diabetes.

Economic versus Spatial Barriers to Healthy Food

One of the most intriguing and interesting results of the analysis is the lack of evidence that the spatial access to healthy and unhealthy food options has a direct association with the odds of diabetes. In fact, a key hypostasis of this study that spatial assess to healthy food is related to the diabetes outcome was not confirmed by the analysis of the available data. One possible reason for the lack of a detectable effect of spatial access on the outcome is that while spatial access limits some of the individual food choices, these limitations are in fact much less pronounced than it was initially assumed. Other factors, such as economic access and culturally driven dietary preferences, discussed later in this chapter, may exert a more powerful effect on the diets and metabolism than the spatial access to stores and restaurants. In fact, the analysis showed that lower neighborhood SES explains a higher probability of having diabetes when the food access is held constant. This finding emphasizes the key effect the neighborhood SES has on diabetes and explains to some degree the lack of measurable spatial effect.

On the other hand, the general spatial pattern of types and densities of food outlets revealed by the mapping of the kernel density estimates was that areas with lower SES also had less total food density and lower access to these food outlets that provided a wider variety of foods, including healthy options. The strong spatial covariance between low SES and low healthy food access found in this study suggests that social and economic factors, may affect the total food context around the residences, including food cost, quality, and availability, and this issue merits a further discussion.

To a varying extent, all residents participate in and shape the socio-economic fabric around their residences, and in turn are affected by it. The size of these effects is impossible to quantify without community survey level data that includes individual-level SES data, and quantitative as well as qualitative measures of the total neighborhood environment. However, literature and common sense both suggest that the degree to which the individuals are affected by the lack of healthy food options around their homes is defined not merely by spatial access, as has been sometimes suggested (Sharkey et al. 2009)--but also by individual and household economic status. This is because in the neighborhoods where spatial access to healthy food is poor and healthier food shopping options lay outside the neighborhood and beyond the walking distance from homes, (such as in "food desert" areas included in this study) individual's economic access plays a critical role in the person's ability to keep a healthy diet. For example, food shopping on the Internet, at a distant supermarket, or at a wholesale food center is probably inaccessible to those with a lower SES without a personal computer and private car but is within the reach of individuals with higher SES who are likely to have such resources. It is important to note that in New York City with its great diversity, the individuals in the former and the latter scenarios may well share the same physical city block or administrative "neighborhood," as defined by census boundaries. This differential economic access creates conditions under which poorer households, which tend to have less access to Internet and are less likely to own a private vehicle, become "captive shoppers," limited to the few unhealthy choices that remain within walking distance from their residences. Due to the limitations of the data available in this study, however, it is not possible to account for these differential levels of access to healthy foods.

Limitations

Building on these observations and on the experience gained in the process of capturing, cleaning, and analyzing large amounts of geo-referenced data related to diabetes, a follow-up study should be designed to concurrently study the neighborhood SES and individual SES effects on food access on diabetes risk. Since diabetes develops over a time and results from long-term exposures to unhealthy food and other neighborhood-level variables, precise longitudinal data on exposures to such contextual variables, hypothesized to be offensive in respect to metabolism, it is not possible to test these linkages in a cross-sectional study. Current research, this study included, has been struggling to respond to this challenge. Specifically, the current study was limited by lack of individual SES data and of subjects' address history. Thus, I could only examine a "slice in the life-time" of the subjects, using only the data collected by clinicians during the ED visit and by appending contextual data based on the spatial location of the current residence. But the neighborhood SES context may change dramatically when subjects move from one area to another. Alas, the EMR data did not provide patient residence address history and it was not possible to analyze such effects. This limitation is especially significant in light of the well-grounded argument that detrimental exposures may have a cumulative effect on the outcome. For example, given the exposures, one may move from being overweight, to being obese, to pre-diabetes, to type 2 diabetes, within a period of a decade.

An interesting corollary finding of this study was that several spatial gaps were identified in respect to food access. Individuals living in these gap areas have very poor spatial access to even most basic types of groceries. It would be methodologically instructive and could provide new insight into energy balance research to study the association between residence in such food deserts and the risk of obesity and diabetes. Clearly, such studies would also require precise data on address history.

Another important aspect of this research that warrants further investigation is the relationship between obesity and diabetes. It would be instructive to repeat this analysis with the outcome of obesity in place of diabetes to derive the odds of subjects of being obese, and to explore to what extent the trends reflect the spatial distribution of these identified food deserts. Unfortunately, the ED from KCMC dataset did not contain height and weight of the subjects, which is needed to compute the body mass index (BMI), required for such additional analysis.

The Challenge of Detecting Undiagnosed Diabetes Cases

One of the key challenges in this and other studies on diabetes is that diabetes is a chronic condition that can have few clinical manifestations and can be easily missed or misdiagnosed in the clinical settings. Its “silent” nature is unlike that of obesity, commonly its concomitant condition. Obesity can be quickly and inexpensively measured by anthropometry alone. But the definitive diagnosis of diabetes relies on the plasma glucose estimation, a relatively expensive test, which remains the basis of clinical diagnostic criteria for diabetes worldwide (WHO 2010). Specifically, the WHO clinical criteria for of the diabetes diagnosis requires the test to be preformed while fasting and its results need to be interpreted by a clinician (Frieden et al. 2008).

However, there are important differences between clinical diagnosis of diabetes to identify an individual with the disease and defining diabetes for epidemiological purposes (WHO 2006). In the former case the diagnosis requires careful substantiation with retesting on another day unless the person is symptomatic and the plasma glucose is unequivocally elevated (WHO 2010). However, the inclusion criteria for an epidemiological study of diabetes are typically much more lenient. For example, repeat glucose testing is rarely performed. Interestingly, when studies do perform repeat testing, only 75% of people with diabetes detected in these epidemiological studies are confirmed to have clinical diabetes (WHO 2006).

Since this study used only clinical data from the EMR, it relied on the more stringent clinical diagnostic requirements, setting it apart from most epidemiologic studies on the subject. However, even under these more stringent case selection conditions, the study was limited by the by the data available from a retrospective review of the patients' EMRs. In clinical settings, there is always a possibility of undiagnosed diabetics passing through the ED. In some situations the clinician's focus may be diverted from diabetes by the nature of the patient encounter. This scenario is most likely when the individuals with new-onset diabetes are seen in the ED for urgent care for other conditions, such as trauma. In these situations no ICD-9 code for diabetes would be included in the EMR. Based on the interviews with clinicians at the ED, the missed diagnosis of diabetes may be especially likely in patients arriving with an acute injury, the most common primary diagnosis for ED admission at KCMC (Table 15 which lists the 10 most common primary diagnoses in KCHC ED visitors during the study).

.

10 most common primary diagnoses in ED visitors during the study (in order of descending frequency)	
ICD9 code	**Description**
959.9	Injury, unspecified site
724.5	Backache, unspecified
784	Symptoms involving head and neck
789.09	Abdominal pain
786.5	Chest pain
729.1	Myalgia and myositis, unspecified
626.6	Metrorrhagia
780.4	Dizziness and giddiness
462	Acute pharyngitis
719.4	Pain in joint
599	Other disorders of urethra and urinary tract
616.1	Vaginitis and vulvovaginitis

Table 15: 10 most common primary diagnoses in ED visitors during the study.

Recognizing the limitations of selecting cases and controls solely by utilizing ICD-9 codes in EMR, I made an additional effort to include all individuals with diabetes in the sample by using both diabetes-specific ICD-9 codes and the WHO's clinical diabetes diagnosis criteria for diabetes diagnosis using HbA1c values, discussed in Chapter 3. I used the American Diabetes Association's clinical cut-off values for HbA1c levels for the diagnosis of diabetes mellitus, which suggest that the diagnosis should be made if the HbA1c level is more or equal to 6.5% (Greiver et al. 2011; Ettner et al. 2009; ADA 2008).

Utilization of HbA1c data permitted the inclusion of several hundred additional diabetes cases. This suggested that a significant number of diabetics that were seen at KCMC ED were not assigned ICD-9 diagnostic codes for diabetes. The finding that a large number of undiagnosed diabetics passed unnoticed through the ED constitutes a significant and important result of this study and deserves a thorough investigation. It points out the relatively low reliability of ICD-9 codes in selecting diabetes cases from the hospital population using EMR and underscores the importance to augment such sample selection process with objective measurements, such as the HbA1c criteria used in this study. One approach to achieving this is by offering free HbA1c testing to all incoming patients in the ED during the study period. At least one group of researchers at Massachusetts General Hospital developed reliable protocols for such a prospective study, which should be considered in future studies (Wexler et al. 2008).

EMR Extracts versus Population-based Data Sources

This study tested the utility of electronic medical records as a source of health data for neighborhood health research. It has been reported that over 80 percent of all health care data have a geographic component (Anonimous 2000). EMR is a new but increasingly important source of geo-referenced public health data. It is a very low-cost source of data and has practically no added participant-burden because the data is being collected during clinical care. The trend of replacing paper medical records with EMRs is also reflected in the policy of the current U.S. Administration, which argues for a complete conversion of all medical records into EMRs in the next five years (Bast and Six 2010). This sweeping change from paper to electronic media for medical record-keeping will potentially make EMRs available robust sources of new health data. However, based on the experience gained from this study, the complex implications of using EMR for research on diabetes and its limitations should also be recognized in future studies. These issues are briefly discussed in this section.

In the US most EMR systems are typically intricate proprietary systems with bundles of database software, data input software (commonly known as EMR), and various hardware. For patient protection reasons health and other personal data collected by these systems are located behind the institutional firewalls of large health care centers. This patient data is typically not shared with community-based health care providers, such as primary care doctors, who practice outside theses medical centers. This condition creates several barriers for data continuity, which is highly problematic for research aiming to study neighborhood effects on health over time. I discuss both of these issues using Kings County Medical Center as a study case.

Health data continuity

The EMR data stored at KCMC is available only to the clinicians that care for the patients within the hospital and does not include individual medical records maintained by community primary care providers, such as family doctors. In absence of a unified EMR policy that affords to electronically share patient data between hospitals and community-based health care providers, there is no practical way for the researchers utilizing a hospital EMR as their primary data source to learn about the true prevalence of diabetes in the community. This is because individuals with diabetes are more likely to be seen and diagnosed outside of the hospital setting, most often by their primary care provider.

Another issue with utilizing the EMR data and neighborhood health is that EMR records often lack address history and thus the length of exposures at each address cannot be calculated. Since it cannot be assumed that study participants resided at the same address or even within the same neighborhood for their entire lives, time and location-specific exposures cannot be calculated. This presents an especially serious limitation in using EMR-derived datasets for longitudinal studies of urban populations in major US cities where large proportion of the local population is foreign-born and is geographically mobile. Lack of information on the length of stay at a particular address makes it impossible to create time and location-stamped record of exposures. Such record is necessary for longitudinal analyses of linkages between neighborhood context and diabetes, which may provide more insight into this complex problem.

Due to the lack of unified electronic record-keeping systems that include diagnostic information from both hospital and community-based health care providers in the US, currently the most accurate information to determine true diabetes prevalence and incidence would stem from a population-based study. Clearly, such comprehensive studies, encompassing samples from large urban populations with several millions of residents, are very expensive to conduct but few alternatives exist.

Data collected through such large, population-based studies may provide new insight into spatial, dietary, behavioral and SES antecedents of the disease. However, most of the current health surveys, including NYC HANES and Behavioral Risk Factors Surveillance System (BRFSS), are administered by phone and do not collect the precise address and address *history* data required (such as the current residence address and previous addresses of the responders). This also presents a critical barrier to studying diabetes longitudinally in context of the fast-changing neighborhood SES and food contexts, which vary greatly from one neighborhood to the next and from one time period to another.

However, comprehensive community-based survey based studies that use individual-level data are costly and not always practical alternatives to using EMR data. The EMR systems have the potential to become useful sources of in health data for research. But this potential can only be realized if the EMR encompasses both a comprehensive medical history of the individual’s care across all providers (both community and hospital-based) and provides patient’s address history. Once the EMR systems achieve such level of sophistication, they will become inexpensive and robust sources of geographically referenced health data for both syndromic surveillance and chronic disease prevention efforts. Currently, as this study demonstrated, these systems still pose many barriers to timely and effective analyses. Data extraction and manipulation requires a concerted and demanding effort among the research team, clinicians and the IT personnel. While currently EMR data can be utilized in cross-sectional studies, the current lack of data on comprehensive medical history, health behaviors and address history all present major barriers for researchers and clinicians alike. These barriers need to be remedied to clear the path to studying neighborhood health to achieve better health through research-informed prevention. Ensuring sufficient privacy safeguards to maintain the patients’ anonymity is both ethically imperative and critical to the success of such efforts.

Limitations in the Assessment of the Neighborhood Environment

This study focused on areas of Northern and Central Brooklyn and East New York, where the problem of diabetes is especially severe. Several limitations of the data on neighborhood environment data were found during the study and require further discussion in order to pave the way for future research.

Built Environment Effects

Some neighborhood environment (e.g., some food and built environment factors) previously reported to affect metabolism were not integrated into the analyses. Similar to the consumption of fast food, low physical activity has been shown to increase risk of diabetes. For physical activity may be associated with the availability of safe and walkable sidewalks and parks, as well as the quality of these resources. Likewise, building design, availability of different types of active and passive transportation and levels of neighborhood crime all can be hypothesized to affect both diets and the levels of physical activity. However, this data was not available for the current study. These variables should be included in future analyses of contextual antecedents of diabetes.

Modeling approach to Estimate Access to Food Environments

To measure access to different types of food outlets, individual walking times required to reach each of the seven types of food outlet for every subject in the sample were calculated. Raster-based statistical surface approach requires considerably more data manipulation and computer processing time than the network buffer approach used in

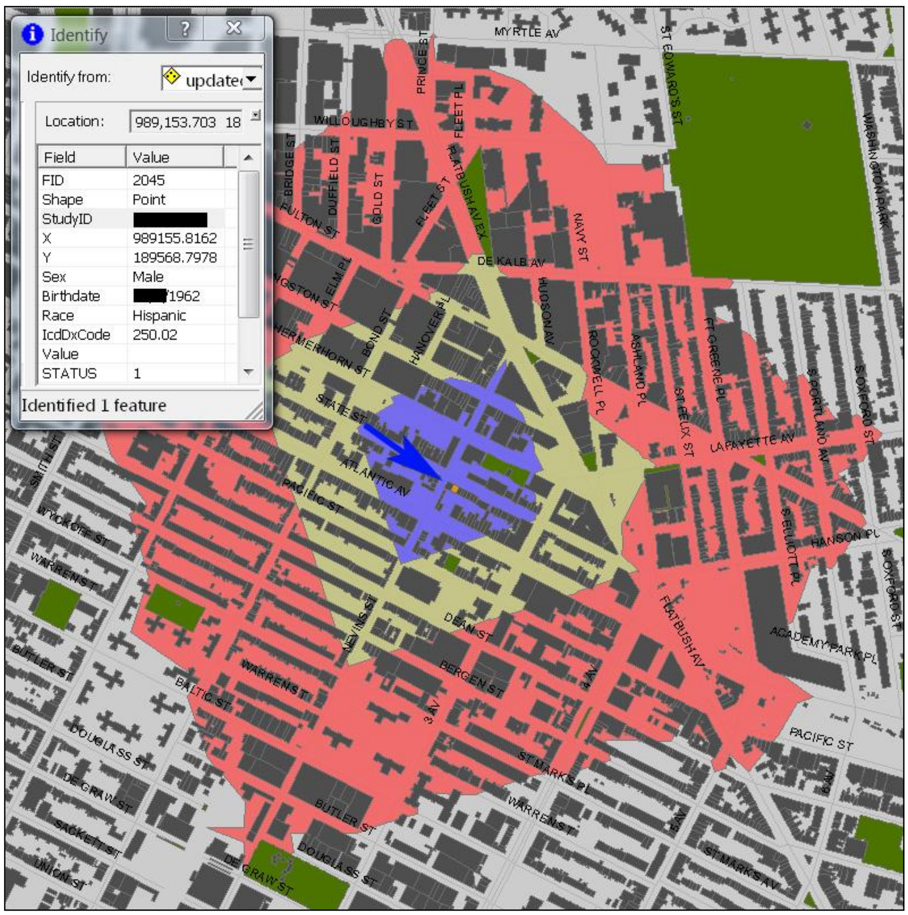

Figure 16. Example of network-analysis based foodshed

earlier studies (see Figure 16, example of network-based foodshed analysis) on food environment. However, because a raster-based model provides a higher precision of estimated travel times, this method has distinctive advantages as compared to the network buffer method used in other studies (see for example Rundle et al 2009) and was thus should be used in future studies where walking access is considered. An additional advantage of the raster-based method is its responsiveness to the critique that food shopping in New York, with its many food shopping options, typically takes place on foot and near where people live (Auchincloss et al. 2008; Khan et al. 2009; Kim et al. 2010; Russell, Hill, and Basser 1998).

While more coarse estimates of food access used in previous studies may be appropriate for other US cities, New York presents a somewhat unique case due to its walkability. In most other large US cities, with their low residential and commercial density, the food outlets are typically located further away from homes. This factor and the higher rates of private car ownership make it possible and often necessary for individuals to travel greater distances for food shopping. But this is not the case in New York City, where private car ownership is relatively low and residents often rely on public transportation or walking to reach food outlets. While there was no data available on car ownership status by the study's subjects, in Brooklyn with its 881,215 housing units, more than half (54.0% or 475,747 units) do not own vehicles (US Census Bureau 2000).

With few exceptions (see Xu et al. 2010) research assessing effects of neighborhood environments on health, including this study, does not control for the heterogeneity of transportation options in different neighborhoods by different individuals. Rather, I assumed that the subjects in the sample typically walk to the closest food stores and restaurants and carry food back to the residences or consume it outside the home and are likely to use the closest stores and restaurants, an effect known as first opportunity shopping (Ettner et al. 2009). Therefore, measuring spatial access to the closest food outlets in the vicinity of the residences was critical to meeting the goal of accurately characterizing the food environment. Another advantage of using raster approach is that it produces a continuous independent variable matrix that can be used in the regression Models to assess the associations between food and diabetes for all of Brooklyn and enables the researcher to examine any selected set of residences in Brooklyn and assign a odds ratios to each area.

A relevant concluding note on the advantage of the raster-derived models relates to the differences in the way the location data is stored and mapped in 3-D space of GIS models. In vector-based schema, non-Euclidean distances are often measured by creating the optimal paths to the destinations to determine the relevant distances between locations. These distances from each location are then used as inputs into the interpolation Model to predict distances that are off of the paths (Ganio et al. 2005). In the literature, such derivation of distances is usually determined by a vector-based solution. The relationship is represented as a network of connected line segments. Although vector-based solutions to deriving non-Euclidean distances for interpolation Models are computationally more efficient than raster based ones, they are of limited utility when non-Euclidean distances cannot be correctly represented by one-dimensional lines. This is the case when modeling steep terrains represented by three-dimensional surfaces, sidewalks that wind between buildings or trees, or paths along the winding, undulating river beds, etc. Therefore, in this study I used a more general raster-based solution to derive non-Euclidean, cost-weighted distances. This technique can be applied to any complex three-dimensional surface.

Lastly, due to the limitations of the food environment data, the contribution of foods purchased online, from mobile street vendors such as the one in Figure 18 and delivered meals could not be accessed in this study. It is not clear how significant the relative contribution of these food vendors is to the overall diets, and this is yet another area that should be investigated in future studies.

Figure 17. A street vendor in central Brooklyn.

Future Research Directions

This study demonstrates the complexity of the relationship between the neighborhood SES, food access, metabolic imbalance and diabetes. While it focused on Brooklyn, the lessons learned from this study are generally applicable to improving neighborhood health elsewhere in urban America.

The relationship between the SES, food access, and diabetes can be deconstructed into two interconnected components. Firstly, low economic status of individuals may affect their odds of diabetes directly because lower-calorie, metabolism-healthy foods, which are protective against diabetes, often cost more per pound (Patterson 2010). Thus, families in poverty may not be able to afford to buy healthy foods and resort to the unhealthy options, which are easily available. Concomitantly, the low neighborhood SES exerts an effect on the types of food outlets that are co-located within the disadvantaged neighborhoods. The foodscape in these areas is dominated by small outlets which primarily sell unhealthy, high energy-density foods, such as corner stores, bodegas, convenience stores, delis, fast food restaurants, and takeout outlets. Such hinged relationship between individual SES and neighborhood SES, and the economics of food shopping is reflective of the complex nature of the relationship between SES, diet and diabetes.

Building on the lessons learned from this study I hope to continue to research the complex relationship between the neighborhood environment and diabetes. An important research direction that has not been investigated here is the study of the relationship between individual-level SES, neighborhood-level SES, and diabetes. Limited by the available data to analyze only the neighborhood SES and food access effects on diabetes I was not able to study this critical component. As a result, the study did not assess to which degree low individual SES has an effect on diabetes when the neighborhood SES is held constant. The findings that the neighborhood SES has an effect on the odds of having diabetes warrant a new study that will also include the individual SES data in the analysis. Such study could be developed building on the preliminary data from the current study and by conducting a survey to capture both the individual SES of the subjects as well as the neighborhood SES.

Importance of the Ethnomedical Model

In the process of collecting preliminary data for this project I conducted informal interviews with fifteen Afro-Caribbean residents of Central Brooklyn that came to a local diabetes clinic located at SUNY Downstate Medical Campus in Central Brooklyn. Based on these interviews the formation of relationships between food and illness is often dichotomized into “bad” and “good," “our food” vs. “American food,” with ethnic Caribbean foods perceived as healthy and local, American foods seen as unhealthy. The patients generally believed diabetes to result from consumption of American "junk foods" (hamburgers, sugar, soft drinks with corn syrup). Some interviewees believed that the disease can be avoided by eating traditional Caribbean foods such as pineapple and ginger, and some “clean” animals. The patients generally recognized dietary linkages to diabetes. However, physical activity as a means of preventing metabolic diseases and decreasing the risk for diabetes was not reported as part of this ethnomedical model.

It is important to emphasize that mapping and modeling the relationship between diabetes prevalence, SES and food outlets in the neighborhood is but the first step to effective diabetes prevention. In order to accurately assess the prevalence of diabetes in the community and to develop a successful intervention program to control and reduce diabetes, a participatory research approach needs to be developed. Using mixed qualitative and quantitative data presentation techniques, diabetes and its causes can be explained to individuals in their socioeconomic and cultural context and through the underlying belief systems that affect related activity and dietary behaviors. This approach may include meetings with residents to learn about the local ethnomedical model and to describe which factors the local residents see as key in affecting the prevalence of diabetes within their neighborhood using an array of mapping and data presentation techniques (Tufte 1997, 2001).

Using this ethnomedical data as well as results of the statistical analysis, an intervention strategy to curb diabetes can be developed. For example, diabetes counseling can be tailored to meet the local ethnomedical model. It may range from coaching individuals in physical activity and relating information about healthy and unhealthy food choices to providing training and support for weight loss and lasting diet changes. Free "fingerstick" blood glucose tests and referrals can also be considered. The local residents in turn can help to identify appropriate community nodes that receive a large number of local visitors that may be receptive to diabetes control and prevention programs. These locations may include parklands, shopping malls, specific street corners and places of worship. In summary, it is critical to carefully include the local ethnomedical models into prospective observational and interventional studies. The local residents should be seen as active participants, fully included in planning and conducting research and prevention programs. This is important to ensure acceptance and local buy-in in participatory research, which is often a critical aspect of such efforts.

This study showed the effectiveness of using spatially-referenced data for outcomes to detect and investigate clustering of diabetes in communities in areas of central eastern and northern Brooklyn to further investigate the causal mechanisms that drive appearance of these clusters and to plan effective interventions. Cluster risk ratios maps can derived and easily from the hospital data by using case and control groups in the Bernoulli model for dichotomous outcomes data. These local cluster scans, such as one shown on Figure 15, compare the relative risk of having diabetes per each area as small as census block. Thus, they can be used directly in diabetes control and prevention efforts. Specifically, in the census blocks with clusters where the RR's are high, as these in the middle of the study area, a participatory strategy for prevention and control of metabolic disease should be developed. A consortium of public health officials, residents, local food outlets, and fitness facilities operators, local farmers and health care providers should be convened to craft and implement a multi-faceted strategy to address all aspects of this complex and devastating and exceedingly expensive health problem. The key goal for such efforts should be to provide a comprehensive prevention and control strategy with measurable outcomes. For example, evidence that healthy food environments are protective against diabetes can be used to improve business opportunities for green grocers and for stores that sell healthy foods with low energy density. Encouraging recreational physical activities and active transportation (biking and walking) can also be effective interventions in curbing the epidemic of diabetes, heart disease and other ailments linked to metabolism.

In Conclusion

Directed by the seminal literature on diabetes, neighborhood health, and spatial analysis, and guided by the advice of Dr. Maantay, and of my doctoral committee members, and by the dedicated clinicians at Kings County Medical Center, I was able to construct SES environment and food access in Brooklyn and statistically quantify their effect on the outcome of diabetes. The strength of cross-sectional design used in this study is in that it measures the probability of a diabetes outcome due to the local factors hypothesized to be associated with diabetes. Its obvious weakness is the lack of ability to measure how duration of exposures to socioeconomic deprivation and poor food access affects an outcome of diabetes using multi-level logistic models. Building on this foundation of knowledge, I plan to expand my research to include both spatial and temporal effects of these exposures on an outcome of diabetes and on other metabolic-spectrum diseases.

References

ADA (2008) Economic costs of diabetes in the U.S. In 2007. *Diabetes Care,* 31**,** 596-615.

Adamo, K. B. & F. Tesson (2008) Gene-environment interaction and the metabolic syndrome. *Novartis Found Symp,* 293**,** 103-19; discussion 119-27.

Althoff, K., A. Karpati, J. Hero & T. Matte (2009) Secular Changes in Mortality Disparities in New York City: A Reexamination. *Journal of Urban Health,* 86**,** 729-744.

Anonimous (2000). *Health Data Management,* 24**,** 24.

Auchincloss, A. H., A. V. Diez Roux, D. G. Brown, C. A. Erdmann & A. G. Bertoni (2008) Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. *Epidemiology,* 19**,** 146-57.

Bader, M. D., M. Purciel, P. Yousefzadeh & K. M. Neckerman (2010) Disparities in neighborhood food environments: implications of measurement strategies. *Econ Geogr,* 86**,** 409-30.

Barron, H. V., A. Rundle, J. Gurwitz & A. Tiefenbrunn (1999) Reperfusion therapy for acute myocardial infarction: observations from the National Registry of Myocardial Infarction 2. *Cardiol Rev,* 7**,** 156-60.

Bast, J. & M. Six (2010) Maintaining patient satisfaction in an EMR conversion. Focus on service and communications is key. *Healthc Exec,* 25**,** 48-50.

Beebe-Dimmer, J. L., N. L. Nock, C. Neslund-Dudas, A. Rundle, C. H. Bock, D. Tang, M. Jankowski & B. A. Rybicki (2009) Racial differences in risk of prostate cancer associated with metabolic syndrome. *Urology,* 74**,** 185-90.

Burch, W. R., N. H. Cheek & L. Taylor. 1972. *Social behavior, natural resources, and the environment*. New York,: Harper & Row.

Chen, Y. Y., S. V. Subramanian, D. Acevedo-Garcia & I. Kawachi (2005) Women's status and depressive symptoms: a multilevel analysis. *Soc Sci Med,* 60**,** 49-60.

Codd, E. F. (1970) A Relational Model of Data for Large Shared Data Banks. *Information Retrieval*.

Congdon, P. (2010) A multilevel Model for comorbid outcomes: obesity and diabetes in the US. *Int J Environ Res Public Health,* 7**,** 333-52.

Curtis (2010) Spatial patterns of diabetes related health problems for vulnerable populations in LosAngeles. *International Journal of Health Geographics,* 9**,** 43.

Daniel, M., S. Moore & Y. Kestens (2008) Framing the biosocial pathways underlying associations between place and cardiometabolic disease. *Health & Place,* 14**,** 117-132.

Darmon, N. D., K (2008) Does social class predict diet quality? *Am J Clin Nutr*.

Deshpande, A. D., E. A. Dodson, I. Gorman & R. C. Brownson (2008) Physical Activity and Diabetes: Opportunities for Prevention Through Policy. *Physical Therapy,* 88**,** 1425-1435.

Diez Roux, A. V. (2009) The persistent puzzle of the geographic patterning of cardiovascular disease. *Prev Med,* 49**,** 133-4.

DiNapoli, P. P. & J. B. Lewis (2008) Understanding school-age obesity: through participatory action research. *MCN Am J Matern Child Nurs,* 33**,** 104-10.

Ershow, A. G., A. Ortega, J. Timothy Baldwin & J. O. Hill (2007) Engineering approaches to energy balance and obesity: opportunities for novel collaborations and research: report of a joint national science foundation and national institutes of health workshop. *J Diabetes Sci Technol,* 1**,** 95-105.

Ettner, S. L., B. L. Cadwell, L. B. Russell, A. Brown, A. J. Karter, M. Safford, C. Mangione, G. Beckles, W. H. Herman, T. J. Thompson & T. S. Group (2009) Investing time in health: do socioeconomically disadvantaged patients spend more or less extra time on diabetes self-care? *Health Econ,* 18**,** 645-63.

Field, D. R. & W. R. Burch. 1991. *Rural sociology and the environment*. New York: Greenwood Press.

Ford, M. M., A. Weise, S. Stokes-Buzzelli, G. Martin & J. Yang (2008) 125: Analysis of the Community Resources for Emergency Department Overuse (CREDO) Initiative for Emergency Department Frequent Users. *Annals of Emergency Medicine,* 52**,** S80-S81.

Frieden, T. R., M. T. Bassett, L. E. Thorpe & T. A. Farley (2008) Public health in New York City, 2002-2007: confronting epidemics of the modern era. *Int J Epidemiol,* 37**,** 966-77.

Friedman, S. M., C. B. Maslow, J. Reibman, P. S. Pillai, R. M. Goldring, M. R. Farfel, S. D. Stellman & K. I. Berger (2011) Case-Control Study of Lung Function in World Trade Center Health Registry Area Residents and Workers. *Am J Respir Crit Care Med*.

Gale, E. A. & K. M. Gillespie (2001) Diabetes and gender. *Diabetologia,* 44**,** 3-15.

Galvez, M. P., K. Morland, C. Raines, J. Kobil, J. Siskind, J. Godbold & B. Brenner (2008) Race and food store availability in an inner-city neighbourhood. *Public Health Nutr,* 11**,** 624-31.

Gangwisch, J. E., S. B. Heymsfield, B. Boden-Albala, R. M. Buijs, F. Kreier, T. G. Pickering, A. G. Rundle, G. K. Zammit & D. Malaspina (2007) Sleep duration as a risk factor for diabetes incidence in a large U.S. sample. *Sleep,* 30**,** 1667-73.

Gehring, U., J. Cyrys, G. Sedlmeir, B. Brunekreef, T. Bellander, P. Fischer, C. P. Bauer, D. Reinhardt, H. E. Wichmann & J. Heinrich (2002) Traffic-related air pollution and respiratory health during the first 2 yrs of life. *Eur Respir J,* 19**,** 690-8.

Goldman, L., C. M. Coussens & National Research Council (U.S.). Board on Health Sciences Policy. Roundtable on Environmental Health Sciences Research and Medicine. 2004. *Environmental health indicators : bridging the chasm of public health and the environment*. Washington, D.C.: National Academies Press.

Gomez-Perez, F. J., C. A. Aguilar-Salinas, P. Almeda-Valdes, D. Cuevas-Ramos, I. Lerman Garber & J. A. Rull (2010) HbA1c for the diagnosis of diabetes mellitus in a developing country. A position article. *Arch Med Res,* 41**,** 302-8.

Greiver, M., B. Aliarzadeh, R. Moineddin, C. Meaney & N. Ivers (2011) Diabetes screening with hemoglobin A1c prior to a change in guideline recommendations: prevalence and patient characteristics. *BMC Fam Pract,* 12**,** 91.

Grove, J. M. & W. R. Burch (1997) A social ecology approach and applications of urban ecosystem and landscape analyses: a case study of Baltimore, Maryland. *Urban Ecosystems,* 1**,** 259-275-275.

Gwynn, R. C., M. Berger, R. K. Garg, E. N. Waddell, R. Philburn & L. E. Thorpe (2011) Measures of adiposity and cardiovascular disease risk factors, new york city health and nutrition examination survey, 2004. *Prev Chronic Dis,* 8**,** A56.

Harper, S. & J. Lynch. 2007. *Selected Comparisons of Measures of Health Disparities: A Review Using Databases Relevant to Healthy People 2010 Cancer-Related Objectives.* Bethesda, MD: NIH Publishing.

Harris, M. I. (1991) Epidemiological correlates of NIDDM in Hispanics, whites, and blacks in the U.S. population. *Diabetes Care,* 14**,** 639-48.

Heraclides, A. M., T. Chandola, D. R. Witte & E. J. Brunner (2011) Work Stress, Obesity and the Risk of Type 2 Diabetes: Gender-Specific Bidirectional Effect in the Whitehall II Study. *Obesity (Silver Spring)*.

Horowitz, C. R., K. A. Colson, P. L. Hebert & K. Lancaster (2004) Barriers to buying healthy foods for people with diabetes: evidence of environmental disparities. *Am J Public Health,* 94**,** 1549-54.

Hotchkiss, A. K., C. V. Rider, C. R. Blystone, V. S. Wilson, P. C. Hartig, G. T. Ankley, P. M. Foster, C. L. Gray & L. E. Gray (2008) Fifteen years after "Wingspread"--environmental endocrine disrupters and human and wildlife health: where we are today and where we need to go. *Toxicol Sci,* 105**,** 235-59.

Hurt, R. T., C. Kulisek, L. A. Buchanan & S. A. McClave (2010) The obesity epidemic: challenges, health initiatives, and implications for gastroenterologists. *Gastroenterol Hepatol (N Y),* 6**,** 780-92.

Jamieson, D. G. (2007) Secondary prevention of ischemic stroke: evolution from a stepwise to a risk stratification approach to care. *Dis Manag,* 10**,** 273-84.

Jemal, A., S. S. Devesa, M. Kulldorff, R. Hayes & J. F. Fraumeni (2000) Geographic variation in prostate cancer mortality rates among white males in the united states. *Annals of Epidemiology,* 10**,** 470-470.

Karnik, A., B. A. Foster, V. Mayer, V. Pratomo, D. McKee, S. Maher, G. Campos & M. Anderson (2011) Food Insecurity and Obesity in New York City Primary Care Clinics. *Med Care*.

Khan, L. K., K. Sobush, D. Keener, K. Goodman, A. Lowry, J. Kakietek & S. Zaro (2009) Recommended community strategies and measurements to prevent obesity in the United States. *MMWR Recomm Rep,* 58**,** 1-26.

Kim, S., K. C. Adamson, D. R. Balfanz, R. C. Brownson, J. L. Wiecha, D. Shepard & W. F. Alles (2010) Development of the Community Healthy Living Index: A tool to foster healthy environments for the prevention of obesity and chronic disease. *Preventive Medicine,* 50**,** S80-S85.

Kirkpatrick, S. I. & V. Tarasuk (2010) Assessing the relevance of neighbourhood characteristics to the household food security of low-income Toronto families. *Public Health Nutr,* 13**,** 1139-48.

Kulldorff, M., C. Song, D. Gregorio, H. Samociuk & L. DeChello (2006) Cancer Map Patterns: Are They Random or Not? *American Journal of Preventive Medicine,* 30**,** S37-S49.

Kulldorff, M., T. Tango & P. J. Park (2003) Power comparisons for disease clustering tests. *Computational Statistics & Data Analysis,* 42**,** 665-684.

Lavin, M. (1998) Comparing America's Most Comprehensive Business Directories. *Database,* 21.

Lee, J. M., M. M. Davis, R. K. Menon & G. L. Freed (2008) Geographic distribution of childhood diabetes and obesity relative to the supply of pediatric endocrinologists in the United States. *J Pediatr,* 152**,** 331-6.

Levine, R. & A. Norenzayan (1999) The pace of life in 31 countries. *Cross-Cultural Psychology,* 30**,** 178-205.

Loftin, C. & S. K. Ward (1983) A spatial autocorrelation Model of the effects of population density on fertility. *Am Sociol Rev,* 48**,** 121-28.

Longley, P. A. & A. D. Singleton (2009) Linking Social Deprivation and Digital Exclusion in England. *Urban Studies,* 46**,** 1275-1298.

Lopez-Zetina, J., H. Lee & R. Friis (2006) The link between obesity and the built environment. Evidence from an ecological analysis of obesity and vehicle miles of travel in California. *Health Place,* 12**,** 656-64.

Maantay, J. (2007) Asthma and air pollution in the Bronx: methodological and data considerations in using GIS for environmental justice and health research. *Health Place,* 13**,** 32-56.

Maantay, J. A., J. Tu & A. R. Maroko (2009) Loose-coupling an air dispersion Model and a geographic information system (GIS) for studying air pollution and asthma in the Bronx, New York City. *Int J Environ Health Res,* 19**,** 59-79.

Maroko, A. R., J. A. Maantay, N. L. Sohler, K. L. Grady & P. S. Arno (2009) The complexities of measuring access to parks and physical activity sites in New York City: a quantitative and qualitative approach. *International Journal of Health Geographics,* 8.

McGinnis, J. M. & W. Foege (2000) Guide to Community Preventive Services: harnessing the science. *Am J Prev Med,* 18**,** 1-2.

McGinnis, J. M. & W. H. Foege (2004) The immediate vs the important. *JAMA,* 291**,** 1263-4.

Michimi, A. & M. C. Wimberly (2010) Associations of supermarket accessibility with obesity and fruit and vegetable consumption in the conterminous United States. *Int J Health Geogr,* 9**,** 49.

MMWR (2010) M*MWR:* Summary of Notifiable Diseases. *Morbidity and Mortality Weekly Report (MMWR)*.

Morabia, A., P. N. Amstislavski, F. E. Mirer, T. M. Amstislavski, H. Eisl, M. S. Wolff & S. B. Markowitz (2009) Air pollution and activity during transportation by car, subway, and walking. *Am J Prev Med,* 37**,** 72-7.

Morabia, A., F. E. Mirer, T. M. Amstislavski, H. M. Eisl, J. Werbe-Fuentes, J. Gorczynski, C. Goranson, M. S. Wolff & S. B. Markowitz (2010) Potential health impact of switching from car to public transportation when commuting to work. *Am J Public Health,* 100**,** 2388-91.

Morland, K. & S. Filomena (2007) Disparities in the availability of fruits and vegetables between racially segregated urban neighbourhoods. *Public Health Nutr,* 10**,** 1481-9.

Morland, K., S. Wing, A. Diez Roux & C. Poole (2002) Neighborhood characteristics associated with the location of food stores and food service places. *Am J Prev Med,* 22**,** 23-9.

Morland, K. B. & K. R. Evenson (2009) Obesity prevalence and the local food environment. *Health Place,* 15**,** 491-5.

Nuvolone, D., R. Della Maggiore, S. Maio, R. Fresco, S. Baldacci, L. Carrozzi, F. Pistelli & G. Viegi (2011) Geographical information system and environmental epidemiology: a cross-sectional spatial analysis of the effects of traffic-related air pollution on population respiratory health. *Environ Health,* 10**,** 12.

Park, Y., K. M. Neckerman, J. Quinn, C. Weiss & A. Rundle (2008) Place of birth, duration of residence, neighborhood immigrant composition and body mass index in New York City. *Int J Behav Nutr Phys Act,* 5**,** 19.

Parsa, H. G., J. Self, D. Njite & T. King (2005) Why Restaurants Fail. *Cornell Hotel and Restaurant Administration Quarterly,* 46**,** 304-322.

Patterson, E. (2010) Dietary energy density as a marker of dietary quality in Swedish children and adolescents: the European Youth Heart Study. *European journal of clinical nutrition,* 64**,** 356-63.

Pedersen, S. D., J. Kang & G. A. Kline (2007) Portion control plate for weight loss in obese patients with type 2 diabetes mellitus: a controlled clinical trial. *Arch Intern Med,* 167**,** 1277-83.

Reilly, J. J., B. Chin, J. Berkowitz, J. Weedon & M. Avitable (2004) Use of a state-wide administrative database in assessing a regional trauma system: the New York city experience. *Journal of the American College of Surgeons,* 198**,** 509-518.

Rundle, A., S. Field, Y. Park, L. Freeman, C. C. Weiss & K. Neckerman (2008) Personal and neighborhood socioeconomic status and indices of neighborhood walk-ability predict body mass index in New York City. *Soc Sci Med,* 67**,** 1951-8.

Rundle, A., K. M. Neckerman, L. Freeman, G. S. Lovasi, M. Purciel, J. Quinn, C. Richards, N. Sircar & C. Weiss (2009) Neighborhood food environment and walkability predict obesity in New York City. *Environ Health Perspect,* 117**,** 442-7.

Rundle, A., A. V. Roux, L. M. Free, D. Miller, K. M. Neckerman & C. C. Weiss (2007) The urban built environment and obesity in New York City: a multilevel analysis. *Am J Health Promot,* 21**,** 326-34.

Rundle, A. G. & A. I. Neugut (2009) Modeling the effects of obesity and weight gain on PSA velocity. *Prostate,* 69**,** 1573-8.

Russell, C., B. Hill & M. Basser (1998) Older people's lives in the inner city: hazardous or rewarding? *Aust N Z J Public Health,* 22**,** 98-106.

Schröder, H. (2007) Protective mechanisms of the Mediterranean diet in obesity and type 2 diabetes. *The Journal of Nutritional Biochemistry,* 18**,** 149-160.

Sharkey, J. R., S. Horel, D. Han & J. C. Huber (2009) Association between neighborhood need and spatial access to food stores and fast food restaurants in neighborhoods of colonias. *Int J Health Geogr,* 8**,** 9.

Stotts Krall, J. & B. Lohse (2009) Interviews with low-income Pennsylvanians verify a need to enhance eating competence. *J Am Diet Assoc,* 109**,** 468-73.

Thorpe, L. E., U. D. Upadhyay, S. Chamany, R. Garg, J. Mandel-Ricci, S. Kellerman, D. K. Berger, T. R. Frieden & C. Gwynn (2009a) Prevalence and Control of Diabetes and Impaired Fasting Glucose in New York City. *Diabetes Care,* 32**,** 57-62.

--- (2009b) Prevalence and control of diabetes and impaired fasting glucose in New York City. *Diabetes Care,* 32**,** 57-62.

Trucco, M. (2009) Gene-environment interaction in type 1 diabetes mellitus. *Endocrinol Nutr,* 56 Suppl 4**,** 56-9.

Tudor-Locke, C., T. Washington, B. Ainsworth & R. Troiano (2009) Linking the American Time Use Survey (ATUS) and the Compendium of Physical Activities: methods and rationale. *J Phys Act Health,* 6**,** 347-53.

Udasin, I., C. Schechter, L. Crowley, A. Sotolongo, M. Gochfeld, B. Luft, J. Moline, D. Harrison & P. Enright (2011) Respiratory symptoms were associated with lower spirometry results during the first examination of WTC responders. *J Occup Environ Med,* 53**,** 49-54.

van den Hazel, P., M. Zuurbier, W. Babisch, A. Bartonova, M. L. Bistrup, G. Bolte, C. Busby, M. Butter, S. Ceccatelli, A. Fucic, W. Hanke, C. Johansson, M. Kohlhuber, M. Leijs, C. Lundqvist, H. Moshammer, R. Naginiene, A. Preece, R. Ronchetti, G. Salines, M. Saunders, G. Schoeters, N. Stilianakis, G. ten Tusscher & J. G. Koppe (2006) Today's epidemics in children: possible relations to environmental pollution and suggested preventive measures. *Acta Paediatr Suppl,* 95**,** 18-25.

Van Wye, G., B. D. Kerker, T. Matte, S. Chamany, D. Eisenhower, T. R. Frieden & L. Thorpe (2008) Obesity and diabetes in New York City, 2002 and 2004. *Prev Chronic Dis,* 5**,** A48.

Weiss, M. J. 1988. *The clustering of America*. New York: Harper & Row.

Wexler, D. J., D. M. Nathan, R. W. Grant, S. Regan, A. L. Van Leuvan & E. Cagliero (2008) Prevalence of elevated hemoglobin A1c among patients admitted to the hospital without a diagnosis of diabetes. *J Clin Endocrinol Metab,* 93**,** 4238-44.

Yiannakoulias, N. (2009) An integrated framework for the geographic surveillance of chronic disease. *Int J Health Geogr,* 8**,** 68.