Title
The Contribution of Food Vendor Availability to Body Mass, Glycemic Control, and Neighborhood Health Disparities among Patients with Type 2 Diabetes

Author
Zhang, Yun

Publication Date
2015

Peer reviewed|Thesis/dissertation

The Contribution of Food Vendor Availability to Body Mass, Glycemic Control, and Neighborhood Health Disparities among Patients with Type 2 Diabetes

By

Yun Zhang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mahasin S. Mujahid, Co-Chair
Professor Barbara A. Laraia, Co-Chair
Professor Rucker C. Johnson

Summer 2015

ABSTRACT

Background:

Diabetes is a diet-sensitive chronic disease with crippling complications and costs.[1] In concert with other factors, the food retail options available in the residential neighborhood may support or hinder diabetes self-management efforts and influence disease outcomes. Consequently, geographic variation in the availability of specific food vendors may contribute to disparities in diabetes self-care between affluent and deprived neighborhoods.[2–4] In recent years, numerous states and localities in the United States have developed financing programs to increase the availability of supermarkets and other fresh food vendors in underserved neighborhoods. However, the health consequence of increasing physical proximity to supermarkets is still unknown. Moreover, no analyses have investigated whether geographic variation in food vendor availability plays a role in explaining neighborhood health disparities. This overarching goal of my dissertation research is to address these gaps in the evidence base.

Methods:

This dissertation is broken out into three separate sets of analyses that are presented in five chapters. Chapter 1 provides an overview of the research and policy context that motivates this research. In Chapter 2, I investigate whether the association between neighborhood deprivation (NDI) and BMI is mediated by the availability of "healthful" and "unhealthful" food vendors, as is commonly assumed. In Chapter 3, I examine whether increasing supermarket proximity is associated with weight reduction among a subsample of patients living in close proximity to twelve new supermarkets. In Chapter 4, I examine the association between changes in neighborhood supermarket presence (supermarket development and closure) and changes in glycemic control (hemoglobin A1c). Finally, in Chapter 5, I summarize study conclusions and recommendations for future research.

Study subjects for each analysis represent different subsamples of adults from the Kaiser Permanente of Northern California (KPNC) Diabetes Registry, a large (>300,000 members) and well-characterized longitudinal cohort of insurance plan members with diabetes mellitus identified from clinical records and survey self-report. The Registry was an ideal data resource for research on the association between food retail change and change in clinical outcomes because all available electronic medical records (including inpatient, outpatient, laboratory and pharmacy records) could be linked with geospatial measures by the member's address of residence.

Discussion:

These analyses will help further our understanding of how neighborhood deprivation "gets under the skin" and will help clarify the role of neighborhood food vendor availability in shaping clinical outcomes. The findings are directly applicable to current policy discussions on the health

impacts of supermarket development in food deserts and may help policymakers evaluate policy options for improving diabetes outcomes.

DEDICATION

This dissertation is dedicated to my partner, Zac, for his unwavering support and to my parents, Sharon and Dapeng, who encouraged me every step of the way.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I'm thankful for my dissertation advisors, Professors Mahasin Mujahid and Barbara Laraia, who gave so much of themselves to ensure my academic and personal well-being. I am incredibly fortunate and grateful to have the both of you in my corner. I am also in debt to Professor Rucker Johnson for inspiring my interest in public policy analysis and for reviewing and shaping my analyses. Thanks also to Professor Maya Petersen for her statistical guidance both in and outside of the classroom. I'm also lucky to have partners at the Kaiser Permanente Division of Research, Dr. Andrew Karter, Howard Moffet, and Margaret Warton, who provided access to unimaginably rich datasets and who offered invaluable advice on research design and manuscript preparation. Lastly, this dissertation was built off of the excellent work products produced by the Laraia Research Group and the Kelly Lab at UC Berkeley. Thanks to Samuel Blanchard, Janelle Downing, Aracely Tamayo, and Bethany Hendrickson for making this work possible.

CHAPTER ONE

Introduction

INTRODUCTION

DIET AND WEIGHT MANAGEMENT FOR TYPE 2 DIABETES

Type 2 diabetes mellitus is an increasingly common chronic disease associated with crippling and costly complications, including heart disease, stroke, blindness, kidney failure, and loss of digits or limbs.[1] The rapid growth in the prevalence of type 2 diabetes mellitus in recent years is alarming. Based on data from 2009-2012, the CDC estimates that approximately 37% of American adults ≥ 20 years of age have diabetes or prediabetes,[2] a 43% increase compared to just six years earlier.[3] As the prevalence of diabetes mellitus climbs, so does the related economic burden. Medical costs and lost productivity attributed to diagnosed diabetes increased 41% from 2007 to 2012.[4]

Sweeping lifestyle changes are needed at a population-level to curb or reverse this rise in the prevalence of diabetes. It is well established that diet and weight management can prevent type 2 diabetes and improve prognosis for those already living with the disease.[5–7] Eating a balanced diet rich in fruits, vegetables, whole grains, and lean proteins improves glycemic control, which can prevent or delay the complications of diabetes.[8–10] Diet adherence is also important for maintaining a healthy weight, which can minimize the risk of complications.[7]

Health education campaigns for diabetes prevention and management are laudable, but for many, having knowledge about nutrition, weight and glycemic control guidelines is not sufficient to initiate or sustain recommended lifestyle changes.[11] Macrosocial changes, including changes to the social, physical and economic environments within which individuals live, may be needed to encourage healthy dietary habits. Policy interventions on upstream determinants of health, such as the food retail environment, may be able to lower barriers to healthy living and help make healthier choices the default option.[12,13]

NEIGHBORHOOD EFFECTS ON HEALTH

Space and place are key dimensions across which health is patterned. Numerous studies document substantial differences in health behaviors, disease risk, and mortality between areas of higher and lower neighborhood socioeconomic status, even after accounting for neighborhood composition.[14–18]

One common measure of neighborhood SES is the neighborhood deprivation index (NDI), a composite score calculated from administrative records for US Census area units.[19] A recent analysis of the Diabetes Study of Northern California (DISTANCE)[20] found that NDI had a significant positive relationship with obesity and cardiometabolic risk factors among adults with diabetes, independent of personal characteristics. Specifically, body mass index (BMI), glycosylated hemoglobin level (A1c), and systolic blood pressure were significantly higher for each increase in quartile of neighborhood deprivation.[21]

While there may be a direct causal link between neighborhood deprivation and health, neighborhood deprivation is generally viewed as a crude proxy for specific neighborhood stressors and buffers that affect health behaviors and health outcomes more directly.[17] A burgeoning body of research seeks to characterize the specific neighborhood features that mediate the observed association between neighborhood deprivation and health, so as to

understand the mechanisms by which neighborhood deprivation "gets under the skin."[22–24] The goal is to identify potentially modifiable social and environmental factors on which to intervene so as to reduce neighborhood-level health disparities.

HEALTHFUL FOOD AVAILABILITY

There is growing interest in understanding whether the availability of healthful foods is equitably distributed across areas and how healthful food availability shapes dietary behaviors related to a range of chronic diseases.[25,26] Past research has defined healthful food availability in myriad ways but most studies have used the presence, count, proximity, or density of supermarkets in an area.[27,28] The predominance of these metrics is partly due to data availability, as commercial lists of food outlet locations over large areas are less costly to obtain than data on the availability of actual food items based on in-store audits.

Supermarkets are large food stores that typically offer a wide selection of fresh produce, meats, dairy products, baked goods, and frozen and stable-shelf foods. Supermarkets are commonly classified as food stores with certain Standard Industrial Classification (SIC) codes that surpass a threshold for annual sales or number of employees. Although supermarkets typically carry a wide range of healthful and unhealthful food items, prior research has generally characterized supermarkets as healthful food outlets because in seminal studies, they were found to offer a greater variety of healthful food items, (e.g. fresh produce and whole grains), at lower prices, compared to convenience stores and small grocery stores.[29,30]

RESEARCH ON HEALTHFUL FOOD RETAIL AND HEALTH OUTCOMES

Neighborhood healthful food availability has been shown to be cross-sectionally associated with residents' dietary patterns and obesity status even after controlling for individual characteristics.[31–37] In light of persistently high levels of residential segregation by race & SES, these results suggest that differential access to healthful foods could be an important contributor to socioeconomic and racial/ethnic diet-related health disparities. However, no analyses to date have investigated whether differential healthful food availability contributes to diet-related health disparities between more and less deprived neighborhoods. Additionally, few longitudinal studies exist and none have been able to validate the assumption that greater supermarket availability will improve neighborhood health outcomes.[38–41]

The first such longitudinal study, a "natural experiment" in Glasgow, Scotland, exploited the opening of a new supermarket in a low-income postal district.[42] The authors, Cummins et al, found no change in fruit or vegetable consumption following the opening of the supermarket comparing the low-income postal district to a nearby district with similar socioeconomic characteristics. Nine years later, Cummins and collaborators conducted a similar evaluation of a new supermarket development in a Philadelphia "food desert" neighborhood that was financed by the Pennsylvania Fresh Food Financing Initiative.[26] The Philadelphia-based study compared dietary quality and BMI changes in the intervention community with outcome changes in a demographically comparable and geographically distinct "food desert" community which did not have supermarket development. The authors found that the intervention store lifted community perceptions of healthful food availability but did not improve dietary quality or BMI.[42] In both the Glasgow and Philadelphia studies, sample sizes were small (Glasgow: $N = 412$; Philadelphia:

N = 656) and the resulting analysis had poor power to detect modest effect sizes. Moreover, in each case, there may have been other concurrent neighborhood changes in the "intervention" district that were not present in the "control" district that may contribute to dietary quality and BMI.

Using a larger sample, Boone-Heinonen et al (2011) followed prospectively over 5,000 young adults in the Coronary Artery Risk Development in Young Adults (CARDIA) study from 1985 to 2000.[39] The authors found no significant association between supermarket proximity and dietary intake, after adjustment for individual-level and neighborhood-level covariates in a fixed effects model. Although this study is one of the most methodologically rigorous analyses to date, the analysis sample included individuals who moved residence over follow-up. Thus, the analysis fails to separate the variation in supermarket availability that is caused by supermarket changes from the variation in supermarket availability that is caused by people moving. This is a source of confounding bias because residents may select into neighborhoods based on factors related to dietary intake.

In a recent analysis of four waves of the Early Childhood Longitudinal Study – Kindergarten cohort (ECLS-K),[40] greater supermarket availability was found to be associated with a modest increase in BMI among children who did not move residence over the course of follow-up. However, the observed increase did not achieve statistical significance. The study design ensured that effect estimates reflected the impact of neighborhood contextual changes, rather than residential relocation. However, random-effects models were used to estimate the impact of supermarket availability on BMI and this is problematic because random-effects models are valid only under strict data assumptions that may not apply in this research context.

LIMITATIONS OF EXISTING RESEARCH

Neighborhoods are dynamic and the factors that impact residential and store location are complex. Researchers investigating the dietary and health impacts of supermarket availability have struggled to untangle the contextual effects of neighborhood attribute from the compositional effect of neighborhood residents because the relationship between neighborhood food availability and the dietary preferences of its residents is likely to be bidirectional. These challenges are discussed in further detail below.

SELECTION OF STORES INTO NEIGHBORHOODS

The nonrandom selection of food stores into neighborhoods poses the first challenge. Bias arising from store selection occurs because food store openings and closures are influenced by market and economic trends. The same economic forces that spur supermarket development and closure can also affect countless other changes in the community (e.g. other retail, crime, employment opportunities, local government services and traffic). These other changes may also affect the health outcome of interest. If data on potential confounding neighborhood factors is not collected, it is difficult to disentangle the influence of supermarket availability from concurrent neighborhood changes.

SELECTION OF RESIDENTS INTO NEIGHBORHOODS

Additionally, individuals may change residences over time, and location choices are influenced by many factors, including personal preferences, health status and financial resources. The nonrandom selection of residents into neighborhoods poses a challenge for estimating the health effects of neighborhood attributes because the same factors that influence residential selection may also affect health outcomes.[43–45] For example, individuals who maintain healthy diets may be more likely to prioritize access to healthy foods in their choice of where to live. Health status itself may influence residential location. For example, financial hardship from a chronic illness may cause an individual to relocate to a lower-income neighborhood with a poorer food environment. For these reasons and others, inadequate control for confounding by measured and unmeasured individual-level characteristics can induce spurious relationships between neighborhood contextual factors and health outcomes.

SHORTCOMINGS OF COMMON ANALYTIC AND STUDY DESIGNS

Due to the selection issues described above, a longitudinal study design is necessary to establish exposure-outcome temporality and control for neighborhood- and individual-level confounders. Furthermore, among the various statistical options for longitudinal analyses, estimators that use each subject as their own comparison, such as fixed effects and first-difference estimators, are better able to address the neighborhood selection issues discussed above compared to estimators that exploit between-subject variation, such as random effects models**.** Fixed effects and first-difference estimators use within-subject change in the exposure and covariates to predict within-subject change in the outcome. Such models automatically control for measured and unmeasured time-invariant individual characteristics such as race, sex and educational attainment. In contrast, random effects models, while commonly employed, are ill-suited for this research purpose because, like cross-sectional analyses, random effects models assume that the neighborhood exposures are uncorrelated with unmeasured individual and neighborhood characteristics. This is equivalent to assuming that the covariate list includes all of the various factors that influence the selection of individuals and stores into neighborhoods.

Fixed effects or first-difference longitudinal analyses control for residential selection based on time-invariant individual-level characteristics but confounding by time-varying individual-level characteristics as well as time-varying and time-invariant neighborhood-level characteristics are still concerns. One strategy to minimize confounding by time-varying individual level characteristics as well as by time-invariant neighborhood-level characteristics is to use a fixed effects or first-difference model on an analysis sample which only retains observations at each subjects' modal address. In this way, residential location is static across observations for each individual. Although excluding observations at non-modal addresses from the analysis sacrifices a degree of external validity and precision, it ensures that any observed variation in supermarket exposure is attributable to changes in neighborhood context, rather than to residential mobility. This distinction is important from a policy perspective because we are interested in the health consequences of changes in supermarket availability due to store changes rather than to residential relocation.

In order to minimize confounding by time-varying neighborhood-level factors, studies should measure and adjust for changes in neighborhood socioeconomic status and specific neighborhood attributes that are theorized to affect the outcome to the extent that it is feasible. Changes in the availability of other retail including physical activity venues and other food retail can potentially impact diet quality, weight, and other health outcomes, and should be included as

model covariates if applicable.

In summary, research on the impact of supermarket availability on diabetes outcomes is underdeveloped. Few longitudinal studies exist and among them, none adequately adjust for neighborhood-level and individual-level confounding influences in their analytic and study designs. Rigorous longitudinal studies are needed to address these research gaps.

POLICY IMPLICATIONS

Research on the health impacts of supermarket availability is needed to inform public policy decisions. There is growing interest in the development of supermarkets as a potential health policy intervention for reducing neighborhood disparities in nutrition, obesity and the risk of chronic disease. Although findings are mixed,[46,47] several studies have documented that there are fewer supermarkets, worse nutritional status, and increased risk of chronic disease (e.g. cardiovascular disease and obesity) in low-income neighborhoods.[33,48,37] Given these findings, the 2008 Farm Bill instructed the United States Department of Agriculture (USDA) to study the prevalence of food deserts: low-income census tracts in which more than a third of residents live more than one mile away from the closest supermarket. [4950] More recently, the First Lady's Let's Move! Initiative committed to eliminating America's food deserts in seven years.[51] The primary tool at the federal level that will be used for this purpose is the Healthy Food Financing Initiative (HFFI), which provides healthy food retailers with grants and loans to assist with the costs of starting, refurbishing, and operating qualifying food stores in areas designated as food deserts.[52] In parallel, local and state governments have developed or are developing similar programs to correct what many perceive as market failures in food retail.[53–55]

Given the amount of public attention and policy action in this arena, it is surprising that very little is known about how changes in supermarket availability influence diet, the prevalence of obesity and risk factors for chronic disease over time. The underlying assumption of these programs is that the greater availability of supermarkets would improve the health of area residents through easier access to affordable healthful foods. While such effects are plausible, supermarkets may similarly increase access to nutritionally poor and calorie-dense food options such as processed snack foods, soda, and prepared meals. Thus, the net health impact of improved access to supermarket food offerings is uncertain and may vary by compositional and contextual characteristics of stores and neighborhoods.

RESEARCH QUESTIONS, AIMS AND RATIONALE

Diabetes-related behaviors and health outcomes, such as diet quality, obesity, and hemoglobin A1c level vary systematically by level of neighborhood deprivation.[18,21] There is enormous interest in supermarket development as a potential policy intervention for reducing neighborhood disparities in nutritional status and prevalence of disease. However, we have little understanding of the role of the food retail environment in explaining differences in nutritional status and the burden of chronic disease by neighborhood. Only a handful of longitudinal investigations of the impact of food retail availability on diet, obesity and other health outcomes exist,[38,40,56] and none adequately accounted for the dynamic processes of residential and store selection into neighborhoods and the potential biases produced by these processes. This research attempts to address these deficits in our understanding of the contextual barriers to the prevention and management of diabetes through the following three aims.

Aim 1. Examine whether the association between neighborhood deprivation and obesity is mediated by healthful and unhealthful food retail outlet density.

Aim 2. Among patients who live within two miles of a new supermarket development, examine whether decreasing distance to nearest supermarket (as a result of new developments) is associated with BMI change.

Aim 3. Investigate how gains and losses in neighborhood supermarket presence within one mile of the home affect residents' glycemic control (measured by glycated hemoglobin level, A1c) for patients with different levels of glycemic control at baseline.

CHAPTER TWO

Does Food Vendor Density Mediate the Association Between Neighborhood Deprivation and BMI?

ABSTRACT

Background:

In previous research, neighborhood deprivation was positively associated with body mass index (BMI) among adults with diabetes. We assessed whether the association between neighborhood deprivation and BMI is attributable, in part, to geographic variation in the availability of healthful and unhealthful food vendors.

Methods:

Subjects were 16,634 participants of the Diabetes Study of Northern California (DISTANCE), a multiethnic cohort of adults living with diabetes. Neighborhood deprivation and healthful (supermarket and produce) and unhealthful (fast food outlets and convenience stores) food vendor kernel density were calculated at each participant's residential block centroid. We estimated the total effect, controlled direct effect, natural direct effect, and natural indirect effect of neighborhood deprivation on BMI. Mediation effects were estimated using G-computation, a maximum likelihood substitution estimator of the G-formula that allows for complex data relationships such as multiple mediators and sequential causal pathways.

Results:

We estimated that if neighborhood deprivation were reduced from the most deprived to the least deprived quartile, average BMI would change by -0.73 units (95% CI -1.05, -0.32); however, we did not detect evidence of mediation by food vendor density. In contrast to previous findings, a simulated reduction in neighborhood deprivation from the most deprived to the least deprived quartile was associated with dramatic declines in both healthful and unhealthful food vendor density.

INTRODUCTION

Neighborhood deprivation indices are composite measures of area socioeconomic status (SES) commonly used in early neighborhood effects research as crude proxies for area-level deprivation and as predictors of health access and outcomes.[17] Much of our understanding of the relevance of place to health comes from these early ecologic and multilevel studies of the relationship between neighborhood deprivation and disease risk.[58] Our previous research found that, independent of personal characteristics, neighborhood deprivation index had a significant positive and monotonic relationship with body mass index (BMI) and cardiometabolic risk factor control among adults with diabetes.[21] Diabetes is a chronic disease influenced by health-related behaviors including diet and exercise, and thus place-based interventions that promote weight loss or simply weight maintenance may improve long term diabetes outcomes.[7]

The pathways through which neighborhood deprivation index affects BMI are not well understood, but the food retail environment has been proposed as an important mediator and has been shown to have strong cross-sectional associations with BMI in both healthy and chronically ill populations.[34,59,60] Our previous analysis of the Diabetes Study of Northern California (DISTANCE) found that among moderate to high-income subjects, greater neighborhood healthful food retail density was associated with lower obesity prevalence.[61] However, no studies to date examine whether and how much geographic variation in food retail density accounts for neighborhood-level socioeconomic disparities in BMI.

Is neighborhood density of retail food outlets a major contributor to the BMI disparities we observed between more- and less-deprived neighborhoods in this population with diabetes? To address this question, we estimated the total effect, the controlled direct effect, the natural direct effect, and the natural indirect effect of neighborhood deprivation index on BMI through food retail density, accounting for sequential impacts on intermediate behavioral variables along the causal chain. To accommodate complex data relationships, we used G-computation, a causal inference technique based on the Rubin Causal Model counterfactual framework.[62,63]

We hypothesized that the effect of neighborhood deprivation on BMI is explained in part by geographic variation in healthful and unhealthful food vendor density among those living with diabetes. Using data from DISTANCE, a well-characterized, multi-ethnic cohort of Californian adults with diabetes, we estimated the share of the total effect of neighborhood deprivation index on BMI that was explained by differences in food retail density between most and least deprived neighborhoods. Additionally, we explored the behavioral mechanisms underlying the direct and indirect mediation pathways. We estimated the effect of a simulated reduction in neighborhood deprivation index on smoking, physical activity, and dietary adherence and incorporated the impact of those subsequent behavioral changes on BMI.

METHODS

STUDY SAMPLE

Kaiser Permanente Northern California (KPNC) is a large, integrated health care delivery system caring for more than 3 million persons who are representative of the San Francisco Bay and Sacramento regional population.[64] The DISTANCE survey was conducted during 2005-2006 in an ethnically stratified random sample of KPNC members in the diabetes registry (n = 40,735) with approximately equal samples sizes among the five largest ethnic groups (African American, Chinese, Filipino, Latino, and White). As described by Moffet et al. (2009), a total of 20,188 people responded to the survey for a response rate of 62% after adjusting for estimated eligibility among non-respondents.[20] Respondents to the short-form survey (n = 2,393), individuals with type 1 diabetes (n = 826), and individuals with extreme BMI values above the 99th percentile or below the first percentile (n = 335) were excluded, leaving a final analytic sample of 16,634.

OUTCOME: BMI

BMI was calculated from electronic records using the first clinical measurement of height and weight recorded in an outpatient visit after the survey date. Self-reported weight and height from the survey was used (n = 1,226) if individuals had no measured weight and height within two years after the survey. BMI was inverse-transformed in regression models to approximate a normal distribution. In sensitivity analyses, we assessed the robustness of findings using alternative body mass outcomes including binary indicators of obesity (BMI≥30) and severe obesity (BMI≥35).

EXPOSURE: NEIGHBORHOOD DEPRIVATION

Neighborhood deprivation index was calculated based on 2000 US Census housing and population data. Eight census-derived variables comprising six domains (income, poverty, housing, education, employment, and occupation) were used to create the index using principal components analysis of 2,250 census tracts in the 19 counties with more than 25 DISTANCE respondents (Cronbach alpha = 0.93).[19,21] We calculated neighborhood deprivation index quartiles (quartile 1 = least deprived; quartile 4 = most deprived), and assigned quartile values to respondents at the census tract level based on 2006 home address data.

MEDIATOR: HEALTHY AND UNHEALTHY FOOD VENDOR DENSITY

Healthful food vendor density was defined as the kernel density of chain supermarkets (including wholesale clubs), large grocery stores (>$2 million in sales), and produce outlets at the census block centroid of each respondent's 2006 residence. Unhealthful food vendor density was defined as the kernel density of fast food outlets and convenience stores at the same address. Vendor locations were obtained from the 2006 InfoUSA commercial food store database as distributed through ESRI Inc.,[65] and food stores were classified by Standard Industrial Codes

(SIC) as well as by key word, chain name recognition, and annual sales. ArcGIS 10.1[66] (ESRI, Redlands, CA) was used to transform the geocoded vendor locations into a smooth kernel density surface using a 1-mile radius buffer and a quadratic function for inverse distance weighting. Quartiles of kernel density surface scores were used in the analysis. The first quartile represents the smallest and the fourth quartile represents the greatest kernel density of food vendors.

INTERMEDIATE HEALTH BEHAVIORS

We accounted for the effect of a simulated neighborhood deprivation index reduction on intermediate health behaviors (current smoking status, physical activity, and diet adherence) on the theorized causal pathway from neighborhood deprivation index to BMI. Current smokers (yes vs. no) were members who reported having ever smoked 100 cigarettes and currently smoking at the survey date. Physical activity (sufficiently active vs. inactive) was measured using the brief version of the International Physical Activity Questionnaire.[67] Diet adherence, a potential mediator of the pathway between food vendor density and BMI, was assessed by two items. Subjects were asked 1) "On how many days out of the last SEVEN DAYS have you followed a healthful eating plan?" and 2) "On average, over the past month, on how many DAYS PER WEEK have you followed your eating plan?" Responses from the two survey items were averaged, and diet adherence was dichotomized as good (≥ 5 days) vs. poor (< 5 days).

COVARIATES

Age and sex were collected from KPNC administrative data as of the date of survey completion. Census tract residential density was obtained from the 2000 Census and included in regression analyses as quartile indicator variables.[68] Other covariates derived from survey responses include: race/ethnicity, marital status, nativity, household size, education, employment status, value of assets, income-to-federal-poverty-ratio, and diabetes-related locus of control. Given that diabetes-specific locus of control has been associated with self-care behavior,[69] we included responses to two questions as proxies for health-related attitudes that may influence neighborhood preference. Locus of control was assessed by respondents' levels of agreement on a 5-point Likert scale with the statements: "What I do has a big effect on my health and I can avoid complications of diabetes" (internal locus of control) and "Good blood sugar control is a matter of luck and my blood sugars will be what they will be" (external locus of control). Both internal and external locus-of-control measures were dichotomized (≤3, >3).

MISSING DATA

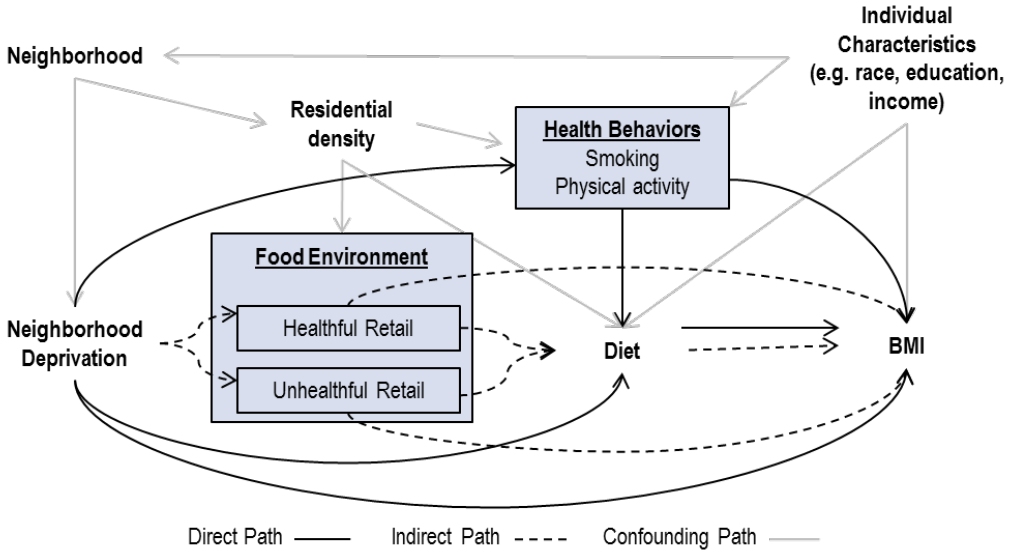
Of the 16,634 respondents in the analytic sample, 7,796 were missing data on at least one of the above measures. Compared to respondents with complete data, respondents with missing data were more likely to be female, non-white, older, and foreign-born, and to have lower income and total years of education. Missing values were imputed using chained equations as described in Appendix Table 1. Since standard errors were estimated using bootstrapping, a

single stochastic imputation was drawn for each missing value in each iteration of the G-computation process.[70]

CAUSAL GRAPH

We developed a directed acyclic graph which makes explicit our assumptions about the causal relationships between neighborhood deprivation index, healthful food vendor density, unhealthful food vendor density, smoking, physical activity, diet adherence, and BMI (Figure 1). In this graph, individuals choose their neighborhood of residence based on preferences, resource constraints, and personal characteristics. Neighborhood deprivation index and residential density are characteristics of the residential neighborhood, and these two factors affect healthful and unhealthful food vendor density as well as individual smoking and physical activity. Healthful and unhealthful food vendor density influences BMI directly and indirectly through diet adherence.

FIGURE 1. DIRECTED ACYCLIC GRAPH.



Arrows represent theorized causal relationships

The decomposition of the effect of neighborhood deprivation index on BMI into direct and indirect effects required adjustment for factors that confound the neighborhood deprivation index-food vendor, food vendor-BMI, and neighborhood deprivation index-BMI relationships. Thus, we adjusted for individual-level traits and residential density in all prediction models. We were also interested in decomposing the natural direct and indirect effect of neighborhood deprivation on BMI into subcomponent pathways that operate through diet adherence. Physical activity and smoking were potential confounders of the diet adherence and BMI relationship; however, these health behaviors were also descendants of the exposure and might mediate the direct effect of neighborhood deprivation on BMI. Standard adjustment for physical activity and smoking could therefore block the direct effect of neighborhood deprivation index on the

outcome. G-computation enabled us to adjust for exposure-dependent confounding by smoking and physical activity without blocking the direct path between neighborhood deprivation index and BMI. The application of G-computation to analyses involving exposure-dependent confounding, of which mediation is a special case, has been discussed previously.[62,70–72] In brief, we estimated the counterfactual probabilities of smoking and physical activity under alternate neighborhood deprivation interventions. Then, to estimate counterfactual diet adherence and BMI values under hypothetical scenarios that intervene on neighborhood deprivation, we set physical activity and smoking to their respective counterfactual values under the specific neighborhood deprivation intervention.

STATISTICAL ANALYSIS

G-computation was used to estimate the total effect, the controlled direct effect, the natural direct effect, and the natural indirect effect of neighborhood deprivation index (A) on BMI using Stata, release 12 (College Station, TX).[73–75] The neighborhood deprivation intervention settings correspond with a hypothetical reduction in neighborhood deprivation from the highest to the lowest quartile. First, we generated prediction models for BMI, healthful food vendor density, unhealthful food vendor density, smoking, and physical activity as a function of their respective variable inputs (Appendix Table 2). Bivariate interaction terms were included as covariates in prediction models if the terms were significant at the $p<0.20$ level. Model diagnostics were performed for all models to identify specification errors. This involved computing the generalized Hosmer–Lemeshow goodness-of-fit statistic[76] for all logistic and multinomial logistic regression models, conducting specification link tests[77] on all models, and visually examining the graph of model residuals against fitted values for BMI.

Next, we used prediction models to simulate counterfactual values for BMI and all variables that impact BMI under six intervention scenarios that differed by exposure and mediator assignment. We simulated counterfactual estimates sequentially such that the predicted probabilities of input variables were used in the prediction of successive outcomes. First, we estimated counterfactual probabilities for each healthful food vendor density quartile (H) and unhealthful food vendor density quartile (U) as well as smoking (S) and physical activity (P) under two exposure settings: 1) if the respondent lived in the least deprived neighborhood [i.e. setting A = 1: to obtain H_{A_1}, U_{A_1}, S_{A_1}, and P_{A_1}], and 2) if the respondent lived in the most deprived neighborhood [i.e. setting A = 4: to obtain H_{A_4}, U_{A_4}, S_{A_4}, and P_{A_4}]. Next, we simulated counterfactual diet adherence probabilities (D) and counterfactual BMI values (transformed back to original units) setting input variables to the counterfactual values outlined in Table 1.

TABLE 1. VALUES FOR VARIABLE INPUTS IN BMI PREDICTION MODELS UNDER SIX INTERVENTION SCENARIOS

Scenario	Neighborhood Deprivation Index Quartile (A)	Smoking status (S)	Physical activity (P)	Healthful Vendor Density (H)	Unhealthful Vendor Density (U)	Diet Adherence (D)
1	1: Least Deprived	S_{A_1}	P_{A_1}	4: Highest	4: Highest	$D\left(A = 1, \mathrm{S} = S_{A_1}, \mathrm{P} = P_{A_1}, H = 4, U = 4\right)$
2	4: Most Deprived	S_{A_4}	P_{A_4}	4: Highest	4: Highest	$D\left(A = 4, \mathrm{S} = S_{A_4}, \mathrm{P} = P_{A_4}, H = 4, U = 4\right)$
3	1: Least Deprived	S_{A_1}	P_{A_1}	H_{A_1}	U_{A_1}	$D\left(A = 1, \mathrm{S} = S_{A_1}, \mathrm{P} = P_{A_1}, H = H_{A_1}, U = U_{A_1}\right)$
4	1: Least Deprived	S_{A_1}	P_{A_1}	H_{A_4}	U_{A_4}	$D\left(A = 1, \mathrm{S} = S_{A_1}, \mathrm{P} = P_{A_1}, H = H_{A_4}, U = U_{A_4}\right)$
5	4: Most Deprived	S_{A_4}	P_{A_4}	H_{A_1}	U_{A_1}	$D\left(A = 4, \mathrm{S} = S_{A_4}, \mathrm{P} = P_{A_4}, H = H_{A_1}, U = U_{A_1}\right)$
6	4: Most Deprived	S_{A_4}	P_{A_4}	H_{A_4}	U_{A_4}	$D\left(A = 4, \mathrm{S} = S_{A_4}, \mathrm{P} = P_{A_4}, H = H_{A_4}, U = U_{A_4}\right)$

The differences in counterfactual BMI values between scenarios 1 and 2 were used to estimate the controlled direct effect of neighborhood deprivation on BMI, holding both healthful and unhealthful food vendor density variables at the highest quartile values. Counterfactual BMI values under scenarios 3, 4, 5, and 6 were used to calculate the total effect, natural direct effect and natural indirect effect of neighborhood deprivation on BMI.

MEDIATION EFFECT PARAMETERS

Assuming that everyone lived in the most deprived neighborhood quartile at baseline, the mediation parameters represent the expected change in population average BMI under alternative neighborhood deprivation and food vendor density interventions. Mediation parameters were estimated as the differences in sample mean simulated BMI comparing the intervention scenarios presented above. Confidence intervals for all mediation effect estimates were obtained from three hundred bootstrap iterations of the G-computation procedure.

CONTROLLED DIRECT EFFECT

$$CDE^{H=4,U=4}$$
$$= E\begin{Bmatrix} Y\left[A=1, S=S_{A_1}, P=P_{A_1}, H=4, U=4, D=D\left(A=1, S=S_{A_1}, P=P_{A_1}, H=4, U=4\right)\right] \\ -Y\left[A=4, S=S_{A_4}, P=P_{A_4}, H=4, U=4, D=D\left(A=4, S=S_{A_4}, P=P_{A_4}, H=4, U=4\right)\right] \end{Bmatrix}$$

The controlled direct effect represents the expected change in population average BMI due to decreasing neighborhood deprivation holding healthful and unhealthful food vendor densities constant at a specific value for all cohort members. Since controlled direct effect parameters are defined by the specific mediator value settings that are chosen, multiple parameter definitions exist. We estimated the controlled direct effect of neighborhood deprivation index on BMI, holding both healthful and unhealthful food vendor density at the highest quartile. This parameter is estimated by the mean difference in simulated BMI for scenario 1 versus scenario 2.

NATURAL DIRECT EFFECT

$$NDE^{H=H_{A_1}, U=U_{A_1}}$$
$$= E\begin{Bmatrix} Y\left[A=1, S=S_{A_1}, P=P_{A_1} H=H_{A_1}, U=U_{A_1}, D=D\left(A=1, S=S_{A_1}, P=P_{A_1}, H=H_{A_1}, U=U_{A_1}\right)\right] \\ -Y\left[A=4, S=S_{A_4}, P=P_{A_4} H=H_{A_1}, U=U_{A_1}, D=D\left(A=4, S=S_{A_4}, P=P_{A_4}, H=H_{A_1}, U=U_{A_1}\right)\right] \end{Bmatrix}$$

Similar to the controlled direct effect, the natural direct effect also estimates the expected change in population average BMI due to decreasing neighborhood deprivation, omitting the influence of food vendor density. However, food vendor density values were set to their counterfactual values under the least deprived neighborhood setting and these counterfactual values may differ between subjects. This parameter was estimated by the mean difference in simulated BMI for scenario 3 versus scenario 5.

NATURAL INDIRECT EFFECT

$$NIE^{A=4} = E\left\{\begin{matrix} Y\left[A=4, S=S_{A_4}, P=P_{A_4}H=H_{A_1}, U=U_{A_1}, D=D\left(A=4, S=S_{A_4}, P=P_{A_4}, H=H_{A_1}, U=U_{A_1}\right)\right] \\ -Y\left[A=4, S=S_{A_4}, P=P_{A_4}H=H_{A_4}, U=U_{A_4}, D=D\left(A=4, S=S_{A_4}, P=P_{A_4}, H=H_{A_4}, U=U_{A_4}\right)\right] \end{matrix}\right\}$$

The natural indirect effect parameter of interest was estimated by the mean difference in simulated BMI for scenario 5 versus scenario 6. This parameter represents the expected change in population average BMI if we could set every study subject to his or her counterfactual food environment values under the least-deprived neighborhood quartile while holding neighborhood deprivation constant at the most deprived quartile.

TOTAL EFFECT

$$TE = E\left\{\begin{matrix} Y\left[A=1, S=S_{A_1}, P=P_{A_1}, H=H_{A_1}, U=U_{A_1}, D=D\left(A=1, S=S_{A_1}, P=P_{A_1}, H=H_{A_1}, U=U_{A_1}\right)\right] \\ -Y\left[A=4, S=S_{A_4}, P=P_{A_4}, H=H_{A_4}, U=U_{A_4}, D=D\left(A=4, S=S_{A_4}, P=P_{A_4}, H=H_{A_4}, U=U_{A_4}\right)\right] \end{matrix}\right\}$$

The total effect (TE) of neighborhood deprivation index on BMI represents the combined effect of all direct and indirect causal pathways. It is estimated by the mean difference in simulated BMI for scenario 3 compared to scenario 6. It is also computationally equal to the sum of the natural indirect effect and the natural direct effect parameters above.

OTHER COMPONENT EFFECTS

The natural direct and indirect effect of neighborhood deprivation index on BMI can be further decomposed into subcomponent pathways through intermediate behavioral outcomes. Appendix Table 3 presents the definitions for these additional subcomponent effects. Each subcomponent effect was estimated as the difference in sample mean simulated values comparing counterfactual interventions that differ by neighborhood deprivation index, healthful food vendor density, and unhealthful food vendor density settings. The estimation process is analogous to the process described above for the estimation of the total effect, natural indirect effect, and natural direct effect.

RESULTS

Table 2 presents selected socio-demographic and health characteristics by neighborhood deprivation index and food vendor density. The average BMI of this cohort was 31.7 kg/m^2. BMI was higher in the most deprived neighborhoods (quartile 4) compared with the least deprived neighborhoods (quartile 1) (32.6 vs. 30.6 kg/m^2). Higher BMI was also observed among residents with the lowest density (quartile 1) compared to the highest density (quartile 4) of healthful food vendors (31.7 vs 31.0 kg/m^2).

TABLE 2. SELECTED HEALTH AND DEMOGRAPHIC CHARACTERISTICS BY NEIGHBORHOOD DEPRIVATION INDEX AND FOOD VENDOR DENSITY

		Neighborhood Deprivation		Healthful Food Vendors		Unhealthful Food Vendors	
	Overall *n*=16,634	Q1: Lowest *n*=3,427	Q4: Highest *n*=3,689	Q1: Lowest *n*=4,148	Q4: Highest *n*=4,144	Q1: Lowest *n*=4,201	Q4: Highest *n*=4,159
BMI (kg/m^2)	31.6	30.6	32.6	31.7	31.0	31.5	31.7
Obese (BMI ≥30)	51%	46%	58%	54%	47%	51%	52%
Severe Obese (BMI≥35)	28%	23%	33%	29%	25%	27%	29%
Follow diet ≥5 days/wk	59%	68%	55%	61%	59%	62%	59%
Physically active	40%	41%	36%	42%	38%	43%	38%
Current smoker	8.4%	6.8%	10%	7.8%	9.5%	7.5%	9.3%
Age	58.6	59.2	58.4	58.5	58.6	58.8	58.5
Percent male	52%	60%	44%	56%	49%	57%	49%
Race							
White	23%	31%	13%	31%	17%	31%	20%
Black	17%	12%	29%	15%	19%	15%	19%
Hispanic	20%	11%	29%	17%	21%	15%	24%
Asian	26%	35%	12%	24%	31%	27%	23%
Other/Mixed	14%	11%	16%	13%	13%	13%	14%
Education							
Less than high school	17%	9.0%	26%	14%	19%	13%	20%
Completed high school or technical	43%	32%	48%	42%	42%	40%	45%
Associates	11%	12%	11%	12%	9.9%	12%	11%
Bachelors	20%	28%	11%	22%	20%	23%	17%
Post-college	9.5%	19%	3.8%	11%	9.4%	12%	7.5%
Married	71%	79%	63%	78%	65%	80%	63%
Income (% of Federal Poverty Level)							
<100%	12%	7.1%	18%	9.2%	14%	8.7%	14%
100-300%	33%	20%	45%	28%	37%	25%	39%
301-600%	35%	36%	30%	37%	32%	37%	33%
600%+	21%	40%	7.3%	26%	17%	30%	14%
Born in USA	64%	61%	67%	66%	55%	65%	60%

Table 3 presents the marginal mediation effect estimates obtained from G-computation. A reduction in neighborhood deprivation from the highest to the lowest quartile would be associated with an average change in BMI of -0.73 units (95% CI -1.05, -0.32). Analyses that used dichotomous outcome classifications (i.e. obesity and severe obesity) found similarly large total effect estimates. We found no evidence of mediation by healthful and unhealthful food vendor density. The simulated effect of lower deprivation on BMI through the healthful and unhealthful retail food pathway was an increase in average BMI of 0.01 (95% CI -0.06, 0.05) units. In sensitivity analyses, healthful and unhealthful food vendor density did not appear to mediate the association between neighborhood deprivation index and dichotomous obesity outcomes.

TABLE 3. MARGINAL MEDIATION EFFECTS OF NEIGHBORHOOD DEPRIVATION REDUCTION FROM HIGHEST TO LOWEST QUARTILE

	BMI (kg/m^2)		Obese* (BMI≥30)		Severe obese* (BMI≥35)	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Total Effect [a]	-0.73	(-1.05, -0.32)	-5.0%	(-7.1, -2.0)	-5.0%	(-6.1, -1.3)
Controlled Direct Effect [b]	-0.71	(-1.00, -0.32)	-4.7%	(-6.8, -1.5)	-4.6%	(-5.6, -0.90)
Natural Direct Effect [c]	-0.74	(-1.06, -0.34)	-6.3%	(-7.0, -1.3)	-5.5%	(-6.8,-0.54)
Natural Indirect Effect [d]	0.01	(-0.06, 0.05)	1.3%	(-0.51, 3.7)	0.5%	(-2.1, 1.9)

* Mediation effects for obese and severe obese dichotomous outcomes are expressed as expected change in probability of event in absolute-scale percentage points.

[a] Total Effect estimates the total effect of a NDI reduction (from highest to lowest quartile) on body mass outcomes.

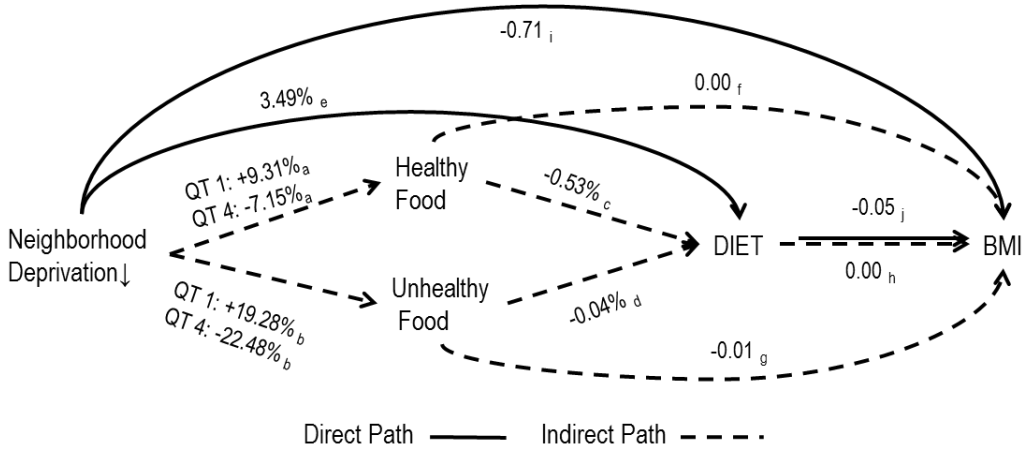
[b] Controlled Direct Effect estimates the effect of a NDI reduction (from highest to lowest quartile) on body mass outcomes holding healthful and unhealthful food vendor density constant at the highest quartile (H=4, U=4)

[c] Natural Direct Effect estimates the effect of a NDI reduction (from highest to lowest quartile) on body mass outcomes, holding healthful and unhealthful food vendor density constant at their counterfactual values under the most affluent neighborhood setting ($H^{A=1}$, $U^{A=1}$).

[d] Natural Indirect Effect estimates the effect of healthful and unhealthful food vendor density change (from $H^{A=4}$ and $U^{A=4}$ to $H^{A=1}$ and $U^{A=1}$) on body mass outcomes, holding neighborhood deprivation constant at the highest quartile.

The natural direct effect and natural indirect effect estimates were further decomposed into constituent subcomponent effects. Figure 2 depicts the simulated influence of neighborhood deprivation index on each intermediate behavioral outcome in the causal chain.

FIGURE 2. DECOMPOSITION OF NATURAL DIRECT AND INDIRECT EFFECTS.



Neighborhood deprivation index was reduced from the most deprived (Qt 4) to least deprived (Qt 1) quartile. Dashed lines indicate path segments that contribute to the natural indirect effect

Unlike many studies that document a positive association between neighborhood affluence and healthful food vendor availability,[48,78,79] in our analysis, a simulated reduction in neighborhood deprivation from the most deprived (quartile 4) to the least deprived quartile (quartile 1) was associated with dramatic declines in both healthful and unhealthful food vendor density. Based on prediction models, the probability of being in the highest quartiles of healthy food vendor density and unhealthy food vendor density would decline 7.2 percentage points (95% CI -11.7, -5.1 and 22.5 percentage points (95% CI: -25.9, -19.6), respectively, if neighborhood deprivation were reduced. The probability of being in the lowest quartiles of healthy food vendor density and unhealthy food vendor density would rise by 9.3 percentage points (95% CI 5.7, 11.6 and 19.3 percentage points (95% CI 17.1, 22.9), respectively, if neighborhood deprivation were reduced. The simulated reduction in neighborhood deprivation index and the resulting changes in healthful and unhealthful vendor density, smoking and physical activity did not appear to improve diet adherence. The total effect of reducing neighborhood deprivation from the highest to the lowest quartile would reduce the probability of diet adherence by 2.9 percentage points (95% CI: -1.4, 5.4).

Simulated BMI and obesity outcomes under the natural course (i.e. no intervention scenario) were comparable to observed outcomes across categories of the exposure and mediators (not shown). Moreover, the magnitude of the deviation between simulated and observed outcome values did not differ across neighborhood deprivation index, healthful food vendor density, and unhealthful food vendor density mediator categories (not shown).

DISCUSSION

In this multi-ethnic cohort of adults living with diabetes, we found a significant association between neighborhood deprivation and BMI using causal modeling methods that adjusted for a wide array of individual characteristics, but we found no evidence that variation in food vendor density mediated that relationship in our sample. Healthful and unhealthful food vendor density accounted for none of the total effect of neighborhood deprivation index on BMI, obesity, and severe obesity.

To our knowledge, this is the first study to evaluate the pathways through which neighborhood deprivation index may affect BMI and obesity using G-computation. Assuming that predictive models are correctly specified, G-computation can estimate natural mediation effect parameters in the presence of exposure-mediator interaction and sequential impacts on intermediate variables along the causal chain. Despite the advantages of using G-computation for mediation analyses and the availability of a statistical package for common mediation parameters (g-formula),[70] few applications of this method have been published to date.[80]

In contrast to previous findings,[48,78,79] greater neighborhood deprivation was associated with a higher density of all types of food vendors, healthful and unhealthful. This is not surprising in this study setting, however. Higher land values in more affluent neighborhoods may discourage construction of supermarkets in these residential areas. Moreover, residents are likely to value the residential feel and, through zoning, maintain a distance from centers of commerce. In addition, produce stores were clustered in neighborhoods with large immigrant populations (i.e. Chinatown, San Francisco), which were less affluent. As expected, unhealthful food venues also clustered in poorer neighborhoods, with the number of unhealthful food venues far outweighing the number of healthful food venues. The associations between food vendor density and neighborhood deprivation index persisted despite adjustment for population density, suggesting that other social forces influence the location of healthful and unhealthful food venues.

Several limitations should be considered when interpreting our study findings. DISTANCE survey respondents represent an insured, managed-care population of adults with type 2 diabetes in a largely urban area, and study results may not generalize to dissimilar populations. Also, given that this mediation analysis is based on cross-sectional data, there is no way to establish time ordering of the exposure, mediators, and outcome. Additionally, the validity of the G-computation estimator hinges on the validity of the predictive models used to create the simulated data. Misspecification of the predictive models, either by omission of confounding variables or miss-specification of functional form would lead to bias. We adjusted for a wide range of individual characteristics and beliefs that plausibly predict residential choice and store location, but these controls may not have been adequate. Moreover, diet adherence, smoking, physical activity, and personal socio-demographic characteristics were obtained from survey responses that may not reflect true values. Additionally, a handful of studies have documented both random and systematic error in vendor classification and identification in commercial databases obtained from marketing firms[81–83] and mediator measurement error can induce bias in mediation effect estimates. Ground-truth verification of supermarket and produce store locations was not feasible given the extensive geographic scope of this study, but store type designations within the commercial list were cleaned and reclassified based on key word searches, name recognition, and annual sales data to improve accuracy.

Overall, retail food vendor density is a crude measure of neighborhood food availability and it is not, on its own, an appropriate measure for neighborhood food access.[27,84,85] In addition to vendor proximity and density, within-store product variety, quality, price, and cultural relevance are important components of food accessibility that may more strongly influence individual food choice. Moreover, resident characteristics such as car-ownership or disability also affect the lived experience of accessing neighborhood food resources. None of these factors were measured in our study.

A related but separate concern that is common to current research on neighborhood health effects is that the exposure and mediator definitions used in analyses do not capture important distinctions in neighborhood environment. Two neighborhoods in the same healthy food vendor density quartile can have vastly different food environments in practice. Additionally, similar food environments can arise from very different underlying processes. For example, a greater density of healthful food vendors may be an unintentional consequence of land use and zoning changes. Alternatively, intentional policy interventions that subsidize food retail development in target neighborhoods may create a similar food environment.[26] Thus, each value of neighborhood deprivation index or food vendor density may in fact represent multiple underlying "versions" of "treatment" that can have differential impact on BMI and this complicates the interpretation of mediation effect estimates.[86,87]

It is also worth noting that there is considerable debate about the policy relevance of natural direct and indirect effects since the intervention settings for the parameters cannot be defined in practice.[88–90] Controlled direct effects, which estimate the magnitude of the exposure-outcome relationship that would remain under an intervention which sets the mediator to specific levels, has been proposed as a preferred alternative.[88,89] However, only natural direct and indirect effects analyses can quantify the relative contribution of a mediation pathway in accounting for the total influence of an exposure on an outcome. In deference to each viewpoint, our study presented both controlled and natural direct effects estimates. We believe that both of these approaches can provide meaningful policy guidance and can help policy makers evaluate and prioritize among many policy options for addressing neighborhood disparities in chronic disease.

We found that neighborhood deprivation is strongly and independently associated with BMI and other weight outcomes, but there was no consistent evidence that these associations were mediated by neighborhood food vendor density in this sample of adults with diabetes. The identification of the specific neighborhood attributes that explain geographic variation in obesity risk in this population remains largely unexplained. Future investigations of the impact of the food vendor environment on diet-related outcomes should adopt a more nuanced exposure definition that makes a distinction between food product availability and retail outlet type. Despite the limitations, our methods and findings contribute to a growing literature which investigates how neighborhood deprivation translates into specific and potentially modifiable neighborhood factors that impact the daily existence of residents.

CHAPTER THREE

Reduction in distance to nearest supermarket was unassociated with BMI change among type 2 diabetes patients

ABSTRACT

Background:

For patients with type 2 diabetes, greater travel distance to a supermarket may hinder healthy eating and weight maintenance. We examined whether a change in supermarket proximity was associated with BMI change among patients who reside within two miles of new supermarket openings.

Methods:

We identified a dozen new supermarkets that opened between 3/14/2009 and 9/15/2010 in eight low- to moderate-income neighborhoods in northern California. Using the Kaiser Permanente Northern California Diabetes Registry, we identified residents with type 2 diabetes in these neighborhoods 12 months prior to the first supermarket opening and ten months following the opening of the last supermarket. Exposure (yes/no) was defined as a reduction in travel distance to the nearest supermarket as a result of a new supermarket opening. First difference regression models were used to estimate the impact of reduced supermarket distance on body mass index (BMI), adjusting for longitudinal changes in patient and neighborhood characteristics.

Results:

The average distance to the nearest existing supermarket was 1.8 miles among the exposed group and 0.8 miles among the unexposed group. Among patients in the exposed group, new supermarket openings reduced travel distance to the nearest supermarket by 0.7 miles on average. However, reduced distance to nearest supermarket was not associated with BMI changes in unadjusted [0.15 (-0.07, 0.40)] or adjusted [0.17 (-0.07, 0.40)] models.

Conclusions:

Overall, we found no evidence that reduced supermarket distance was associated with reduced levels of obesity for residents with type 2 diabetes.

INTRODUCTION

Weight management is an important component of type 2 diabetes disease management but many patients struggle to follow self-care guidelines. In concert with other factors, the food retail options available in an individual's residential neighborhood may shape daily dietary choices and influence body weight and disease outcomes. In recognition of the influence of neighborhood factors on obesity, a growing number of localities and states are developing programs to increase the availability of healthful foods in low income areas.[53,55,91–93] These programs may provide financial incentives for supermarket operators and other food retailers to open or expand in areas with poor supermarket access.

In the epidemiologic literature on the health impacts of the food retail environment, supermarkets are widely regarded as community health assets, because they tend to offer a greater variety of healthful food options (e.g. produce, whole wheat products, lean meats and low-fat dairy) at lower prices compared to smaller food stores.[29,30] However, unhealthful foods such as fried snacks, sugary beverages, and calorie-dense convenience meals may also be more affordable and accessible in supermarkets.

Given the mixed nutritional quality of products at the average supermarket, the net health impact of greater supermarket proximity on body weight is unclear. While early cross-sectional studies found that greater supermarket proximity and density were associated with improved dietary quality and reduced obesity risk,[31–37] very few studies have assessed the health consequences of supermarket development longitudinally. [38–41] In contrast to earlier research, preliminary findings from longitudinal studies do not support the premise that supermarket availability has a beneficial impact on health.

A recent pilot evaluation of a supermarket intervention supported by the Pennsylvania Fresh Food Financing program found that the intervention store lifted community perceptions of healthful food availability but did not improve dietary quality or BMI.[42] This study compared dietary quality and BMI changes in the intervention community with outcome changes in a demographically comparable community which did not have supermarket development. The authors urged other researchers to conduct similar analyses in other locations. However, the logistical challenge of anticipating supermarket developments and collecting longitudinal health outcomes from area residents is a major barrier to study replication.

In contrast to the prospective approach detailed above, we conducted a secondary analysis of medical records in the Kaiser Permanente of Northern California (KPNC) Diabetes Registry. This retrospective approach sidestepped the need to anticipate future supermarket developments and collect original survey responses. Instead, we exploited a dozen previous supermarket developments in eight Northern California neighborhoods as a quasi-experiment to estimate the influence of increasing supermarket proximity on change in BMI among neighborhood residents with diabetes. However, study neighborhoods in the present analysis differ from those in the Cummins et al analysis in a notable way. In the present analysis, although study neighborhood residents lived more than one mile from the nearest supermarket on average, only one of the study neighborhoods qualified as a food desert by USDA criteria.

METHODS

We identified eight neighborhoods (defined below) in northern California in which 12 new supermarkets opened between 3/14/2009 and 9/15/2010. Seven of the 8 neighborhoods had

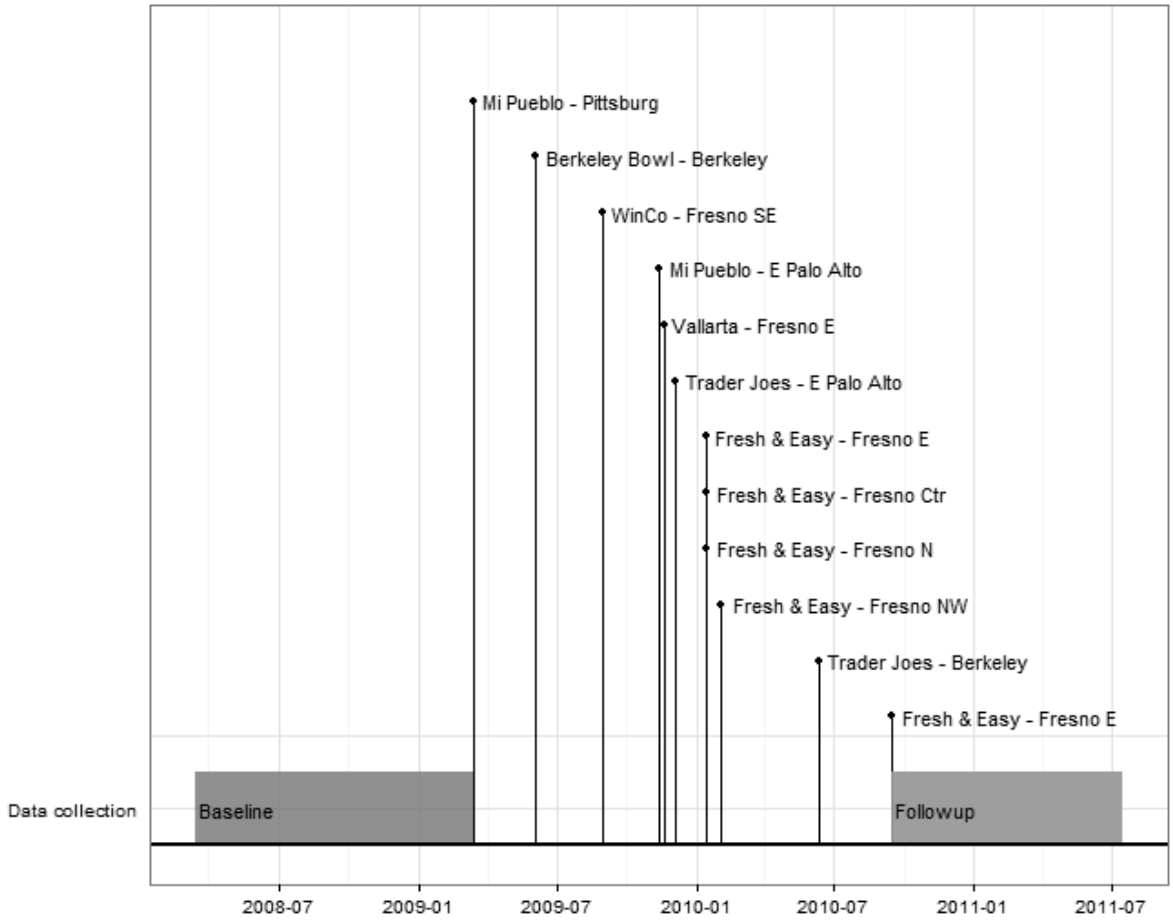
≥15% of residents living below the federal poverty line, and three neighborhoods had ≥20% of residents living in poverty. Only one neighborhood, East Palo Alto, was located in a "food desert" as defined by USDA criteria.[50] There were no other large food retail changes within two miles of neighborhood boundaries in the 12 months preceding (baseline) or the 10 months following (follow-up) this period.

We identified diabetes patients from the KPNC Diabetes Registry (described below) who resided in these eight neighborhoods during baseline and follow-up. Patients who moved residence between baseline and follow-up were excluded. We used geospatial measures to calculate travel distance to the nearest supermarket before and after supermarket openings. New supermarkets were generally located within two miles of existing food stores. Despite this, travel distance to the nearest supermarket was reduced for a subset of residents. Joining geospatial measures with clinical records, we compared BMI changes from baseline to follow-up for patients in each neighborhood who experienced reductions in supermarket distance versus patients in the same neighborhood who did not experience reduced supermarket distance.

STUDY PERIOD

The baseline assessment period (03/14/2008 – 03/14/2009) encompassed the twelve months preceding the first supermarket opening. The follow-up assessment period (09/16/2010 – 07/15/2011) spanned the ten months immediately after the last supermarket opening (Figure 1). The baseline and follow-up assessment period boundaries were constrained to a window of time in which no other large food retail changes (e.g., other supermarket developments, supermarket closures or supercenter grocery expansions) occurred within two miles of the neighborhood boundaries.

FIGURE 1. TIMELINE OF SUPERMARKET DEVELOPMENTS IN STUDY PERIOD



NEIGHBORHOOD DEFINITION

Neighborhood geographic boundaries were defined by a two network mile buffer around each supermarket (Table 1).[1] Overlapping buffer regions were merged into one neighborhood.

[1] In East Palo Alto, neighborhood boundaries were defined as the two network mile buffer around Mi Pueblo only. The region around Trader Joe's was excluded due to its close proximity with a concurrent supermarket closure.

TABLE 1. SUPERMARKET DEVELOPMENTS BY NEIGHBORHOOD

Neighborhood	N	$N_{exposed}$ (%)	Total new openings	Supermarket brands	Days to follow-up[a]	Poverty rate[b]
Berkeley	957	147 (15%)	2	Berkeley Bowl, Trader Joe's	468	15.3%
Pittsburg	677	325 (48%)	1	Mi Pueblo	550	16.3%
East Palo Alto	319	288 (90%)	2	Mi Pueblo, Trader Joe's	304	15.4%
Fresno Northwest	264	33 (12%)	1	Fresh & Easy	224	14.8%
Fresno North	234	38 (16%)	1	Fresh & Easy	245	8.7%
Fresno Southeast	197	35 (18%)	1	WinCo Foods	380	30.0%
Fresno East	461	143 (31%)	3	Vallarta Supermarket, Fresh & Easy (2)	300	20.9%
Fresno Center	138	16 (12%)	1	Fresh & Easy	240	36.0%

[a] Number of days from earliest supermarket opening to first day of follow-up period
[b] Percent of individuals with household income below the federal poverty level (American Community Survey, 2006-2010)

STORE VALIDATION

We defined supermarkets as large grocery stores, supermarkets, wholesale stores, and supercenters with ≥$2 million dollars in annual sales, ≥ 4 cashiers, ≥ 30 varieties of fresh fruit, ≥ 50 varieties of fresh vegetables, ≥ 10 varieties of dairy, ≥ 20 varieties of meat or fish, and ≥ 20 varieties of frozen foods.

Between 05/09/2013 and 07/31/2013, the lead author visited each intervention supermarket to verify store address and confirm that in-store attributes met supermarket criteria. This "ground-truthing" also included verification that, within the study period, no other supermarket developments, closures, or supercenter expansions occurred within a two network mile radius of the intervention neighborhoods. Using 2008, 2009, 2010, and 2011 commercial lists obtained from InfoUSA Inc,[94] a proprietary information service offering commercial databases on business, we identified all supermarkets, supercenters, and large grocery stores within a two mile buffer around neighborhood boundaries. Using official store websites, news articles, Yelp.com,[95]

and Google Streetview,[96] we identified mis-categorized listings and omissions from the InfoUSA list. The dates of store openings and, if applicable, closings for intervention stores and originally existing stores were obtained using the above resources and were cross-checked using off-sale license records from the California Department of Alcoholic Beverage Control[97] and inspection records from the local health department. We contacted supermarket staff to obtain missing information and conducted in-store audits to verify supermarket status of unrecognized store brands.

STUDY SAMPLE

Kaiser Permanente Northern California (KPNC) is a large, integrated health care delivery system whose membership is broadly representative of the overall population in northern California, with the exception of the extreme tails of the income distribution.[64] The KPNC Diabetes Registry is a longitudinal cohort of KPNC members identified with diabetes using clinical records.[98] Individual-level administrative and clinical records for KPNC Diabetes Registry members were linked with geospatial measures by the member's address of residence from membership files.

The study sample is the subset of type 2 diabetes patients from the KPNC Registry who lived within the eight neighborhood boundaries from baseline to follow-up. Members who relocated between baseline and follow-up (n = 385) were removed from the cohort. Members who were pregnant (n = 9) or who had bariatric surgery (n = 10), cancer (n = 175), or lower extremity amputations (n = 24) during or immediately preceding the study period were excluded from analyses due to the impact on changes in weight. Members with missing outcome or covariate values at baseline or follow-up were retained in analyses and variable values were imputed. The final analytical sample was comprised of 3,247 individuals.

OUTCOMES

For each patient, we obtained body mass index (BMI) values from the electronic medical records. Baseline and follow-up average BMI was calculated as the averages of all patient BMI records available in the two respective time periods. The outcome measure was the patient-specific difference in average BMI between the two periods. Patients who did not have a BMI record in the baseline or follow up period were retained and their missing BMI value was imputed.

EXPOSURE

Change in road network distance between each individual's residential census block centroid and the nearest supermarket was calculated as the travel distance in the baseline period minus the distance in the follow-up period using ArcGIS 10.1.[66] The exposure was a dichotomous indicator of a reduction in distance >0.1 mile to nearest supermarket. Given that minor changes in supermarket distance may arise from random geocoding error, the exposed group comprised of residents who experienced a reduction in distance to nearest supermarket by 0.1 mile or more

and the comparison group comprised of residents who experienced less than 0.1 mile reduction in distance.

COVARIATES

Annual clinical measures were extracted from medical and pharmacy records and covariate measures were defined to represent change from baseline to follow-up. The Charlson comorbidity index is a validated measure of mortality risk based on diagnosis of 22 health conditions.[99,100] We also developed separate indicators for medication use associated with weight gain and medication use associated with weight loss based on pharmacy records. Specification of time-invariant individual-level covariates such as gender or race/ethnicity was unnecessary because the statistical model adjusts for these factors.

We adjusted for a number of potential, time-varying neighborhood-level confounders and included indicators for each of the eight neighborhoods in the regression model. Specifically, we adjusted for changes in census block group population density using the difference in five-year aggregate measures from the American Community Survey 2006-2010 to 2008-2012.[101] Additionally, we adjusted for concurrent changes in other retail density including fast food outlet count, produce store count, convenience store count, and physical activity resource kernel density within one network mile of each patient's residential block centroid. For the calendar years associated with the baseline (2008) and follow-up (2011) period, we identified retail locations based on NAICS and SIC industry code, keyword searches, and name recognition in the InfoUSA commercial database for each retail class.

MISSING VARIABLE IMPUTATION

A quarter of the sample were missing values for BMI at baseline or follow-up. We imputed missing values for BMI and Charlson Comorbidity index with predictive mean matching and checked the distribution of imputed values against observed values. Standard errors for model estimates were obtained through bootstrapping of both the imputation and modeling process. In each bootstrap iteration, each missing outcome and covariate value was imputed with a single value using predictive mean matching. To assess the impact of data imputation on model results, we also replicated the analysis using only patients with complete data on all measures (n = 1,908) in a sensitivity analysis.

STATISTICAL ANALYSIS

A first difference regression model was used to model change in individual-level BMI from baseline to follow-up as a function of exposure and change in covariates (Model 1).[102] In a first difference model, the difference in outcome comparing time one to time two is modeled as a function of the differences in exposure and covariate values comparing time one to time two. With two time-periods, the first difference model specification is algebraically equivalent to the potentially more familiar fixed effects model with individual-level fixed effects.

A first difference model offers several advantages for estimating the health impact of non-randomized neighborhood-level interventions. Firstly, covariates that do not change in value across time, such as race/ethnicity or sex, are assumed to have no impact on the direction or magnitude of outcome changes and can be omitted from the model. Additionally, the use of a first difference model combined with sample restriction to residents who remained at the same address over the study period ensured that individual-level changes in supermarket proximity over time were attributable to supermarket developments and not to residential mobility or other individual-level factors such as income or food preferences. In other words, we assumed that changes in a resident's individual-level attributes (ex. personal changes in food preference) have no causal impact on neighborhood supermarket openings or closures and therefore do not affect their change in supermarket proximity.

Model 1:

$$BMI_{it_1} - BMI_{it_0} = \beta_0 + I[(Dist_{it_1} - Dist_{it_0}) < -0.1]\,\beta_1 + (w_{it} - w_{it-1})\beta_2 + (C_i)\beta_3 + (\varepsilon_{it} - \varepsilon_{it-1})$$

In which *i* subscripts the individual, *t* subscripts the time period (0 baseline, 1 follow-up)
$I[(Dist_{it_1} - Dist_{it_0}) < -0.1]$ indicates that travel distance to the nearest supermarket for subject *i* decreased by more than 0.1 mile from baseline to follow-up.
w_{it} is a vector of time-varying covariates including Charlson comorbidity index, medication use associated with weight-gain, medication use associated with weight-loss, PA density, fast food outlet count, convenience store count, produce store count, and population density
C_i is a vector of dummy variables for each neighborhood.

All statistical analyses were conducted in the R programming language.[103] We calculated confidence intervals empirically using 1,000 bootstrap iterations of the data imputation and model fitting process to account for a potentially correlated error structure.

RESULTS

Patients ranged from 23 to 101 years of age with an average age of 64 years (SD=14.7). The sample was ethnically diverse with 35% non-Hispanic white, 24% Black, 22% Hispanic, 10% Asian, and 9% other race (Table 2). The average BMI at baseline was 32.2 (kg/m^2) and over half (56%) of patients were obese. Nearly all patients lived in urban or suburban areas (only ~1% resided in a rural area).

TABLE 2. CLINICAL AND NEIGHBORHOOD CHARACTERISTICS OF STUDY SAMPLE

	Baseline	Follow-up	Difference
	Mean (SD)	Mean (SD)	Mean (SD)
Individual-level variables			
Race/ethnicity			
Non-Hispanic white	35.4%	N/A	N/A
Black	23.7%	N/A	N/A
Hispanic	22.1%	N/A	N/A
Asian	9.6%	N/A	N/A
Other race	9.1%	N/A	N/A
Female	48.0%	N/A	N/A
Age	64.4 (14.7)	N/A	N/A
BMI (km/m^2)	32.2 (7.2)	32.0 (7.4)	-0.2 (2.4)
Comorbidity Score	2.1 (1.5)	2.4 (1.6)	0.2 (1.2)
% on meds associated with weight gain	5.0%	5.9%	0.9%
% on meds associated with weight loss	6.4%	6.7%	0.3%
Area-level variables			
Produce store count (1 mi network radius)	0.4 (0.6)	0.4 (0.6)	0.0 (0.6)
Fast food count (1 mi network radius)	4.0 (3.3)	4.4 (3.6)	0.4 (1.2)
Convenience store count (1 mi network radius)	2.7 (2.1)	2.5 (2.2)	-0.2 (1.2)
Physical activity venue density (1 network radius)	2.0 (1.7)	2.2 (1.9)	0.1 (0.6)
Population density (pop/sq mi) (census block group)	9,432 (5,529)	9,752 (5,806)	320 (2,145)
Federal poverty rate (census block group)	17.6% (12.9%)	19.7% (13.3%)	2.1% (9.6%)

As a result of the new supermarkets opening, a third of patients (32%) were now closer to a supermarket during the follow-up period than in the baseline period. Prior to the opening of the

new supermarkets, the average distance to the closest existing supermarket was 1.8 miles among the exposed and 0.8 miles among the comparison group. Among patients in the exposed group, the new supermarkets reduced supermarket travel distance by 0.7 miles on average (median: 0.6, interquartile range: [0.3, 1.0]). By definition, patients in the unexposed group experienced no change in supermarket distance. The percentage of patients in the exposed group varied across the eight neighborhoods (Table 1). In East Palo Alto, the new supermarket developments decreased travel distance for 90% of patients, with an average of 1.3 mile reduction in distance to the nearest supermarket. In contrast, in Fresno Center, where the new supermarket opened 0.7 miles from an existing store, only 12% of patients experienced a reduction in travel distance to nearest store (-0.3 mile average reduction).

A reduction in the distance to nearest supermarket was not associated with BMI changes in either unadjusted [0.15 (-0.07, 0.40)] or adjusted [0.17 (-0.07, 0.40)] models (Table 3). Neighborhood specific estimates were highly variable (-0.60 to 0.57 BMI units). The data imputation process did not appear to bias study findings. In a sensitivity analysis on a subset of residents with complete measures (n =1,908), the adjusted association between BMI change and reduction in supermarket distance [(0.13 (-0.11,0.39)] was comparable to the estimate obtained from the full sample.

TABLE 3. ADJUSTED ASSOCIATIONS BETWEEN REDUCTION IN SUPERMARKET TRAVEL DISTANCE AND BMI CHANGE

	N	Mean change in supermarket distance among exposed (mi)	ΔBMI	95% CI
Overall				
Unadjusted	3,247	-0.7	0.15	[-0.07, 0.40]
Fully Adjusted[a]	3,247	-0.7	0.17	[-0.07, 0.40]
By Catchment Area[a]				
Berkeley	957	-0.3	0.41	[-0.05, 0.91]
Pittsburg	677	-0.5	-0.22	[-0.60, 0.34]
East Palo Alto	319	-1.4	0.04	[-0.63, 0.94]
Fresno Northwest	264	-0.5	0.57	[-0.04, 1.38]
Fresno North	234	-0.4	0.23	[-0.62, 0.71]
Fresno Southeast	197	-0.5	0.11	[-1.26, 0.49]
Fresno East	461	-0.6	0.33	[-0.11, 0.88]
Fresno Center	138	-0.3	-0.60	[-2.46, 0.85]

[a] Adjusted for Charlson Comorbidity Index, medications associated with weight gain or loss, produce fast food and convenience store count (within 1 mile), physical activity venue kernel density (within 1 mile), and population density

DISCUSSION

In this longitudinal study, we estimated the effect of a reduction in the distance to the closest supermarket on BMI change among patients with type 2 diabetes. We took advantage of the opening of a dozen new supermarkets in eight northern California neighborhoods (March 2009 – Sept 2010) as a quasi-experiment and compared BMI changes over time for patients with reduction in supermarket distance versus patients in the same neighborhoods with no change in distance to nearest supermarket. In most cases, both exposed and unexposed residents lived within one mile of existing supermarket retail and thus, the results of this study are not directly comparable to the aforementioned community health evaluation of the first full service supermarket in a food desert.[42]

Rather, this study estimated the impact on BMI of any reduction in supermarket distance among residents with type 2 diabetes. Contrary to our hypothesis, reduction in travel distance to closest supermarket was not associated with significant or substantive changes in BMI. These findings however are consistent with recent studies which found that the introduction of a full-

service supermarket in a low-income urban neighborhood had no detectable impact on weight or diet-related outcomes among area residents.[41,42]

Our study may, however, underestimate the impact on BMI of reduction in distance to closest supermarket. First, the average reduction in supermarket distance, among patients who experienced any reduction, was small (only 0.7 miles.) While this change may alter food shopping patterns for patients who walk, bicycle, or take public transport to the supermarket, it may not be sufficient behavior change incentive for patients who drive. Second, while we expected the greatest impacts of supermarket openings on patients with shortened distance to nearest supermarket; we anticipated spill-over effects on patients in the comparison group as well, leaving only a modest difference in supermarket exposure between exposure and comparison groups. Lastly, our estimates may understate the impact of supermarket openings on BMI because consumers need time to adjust food purchasing and consumption in response to new food retail opportunities. The time interval between store development and assessment may not have been sufficient to observe the full impact of reduced supermarket distance on BMI outcomes.

Our study has other limitations in addition to the above. This study had limited power to detect small effects in BMI change due to small sample sizes and limited variation in the exposure variable. Also, the time window of this analysis represents the worst recession of the US economy since the Great Depression and our findings may not generalize to different time periods. The supermarket openings in this period of economic recession may be intrinsically different from stores that open in non-recession years; store product mix, prices, marketing, and food purchasing patterns among consumers may have been particular to this time period. Moreover, the analysis sample represents an insured population of patients with type 2 diabetes who were able to remain at the same residence over the study period. Our findings may not be applicable to dissimilar populations. Additionally, we adjusted for localized changes in neighborhood attributes using census block group and one mile radius buffers as granular geographic units. Nonetheless, omitted or inaccurately measured neighborhood-level changes could bias study findings in unexpected directions. Lastly, many factors influence shoppers' choice of food retail outlet aside from physical proximity. These other store qualities, such as price, quality, variety, cultural fit, cleanliness, safety, and customer service may affect patients' visits to and purchases at new supermarkets. While we verified that all 12 intervention stores stocked a wide variety of fresh produce, meats, dairy and frozen items, we did not conduct comprehensive in-store audits, nor did we verify whether patients frequented intervention stores in the follow-up period.

In summary, our study corroborates other recent longitudinal analyses which failed to detect a beneficial effect of increasing residential proximity to supermarkets on diet, weight, and associated health outcomes. Given that supermarkets offer a wide array of both healthful and unhealthful food options; it is time to re-evaluate common assumptions about the impact of their location on health. That said, our findings should not be interpreted as a broad dismissal of all programs that incentivize fresh food retail development in underserved communities. After all, we estimated an average effect of a relatively small reduction in supermarket distance across a wide variety of supermarket types, neighborhoods, and residents. The anticipated effects of improved supermarket proximity on weight outcomes will likely depend on an interaction of resident, store and neighborhood characteristics.

CHAPTER FOUR

Longitudinal associations between neighborhood supermarket presence and hemoglobin A1c among patients with type 2 diabetes

ABSTRACT

Background:

Findings from cross-sectional studies on the influence of food retail on health have shaped the public perception of supermarkets as community health assets. However, few longitudinal studies exist to help policy makers understand the health impacts of supermarket development and closure. We examined whether annual changes in hemoglobin A1c (A1c) were associated with changes in neighborhood supermarket presence among patients with type 2 diabetes.

Methods:

Annual clinical measures (2007 - 2010) from the Kaiser Permanente Northern California Diabetes Registry (n = 434,806 person-years) were linked with annual GIS metrics at each patient's address of longest residence. We specified first difference regression models to estimate the associations between changes in supermarket presence (gain, loss, or no change) and A1c change, adjusting for individual and area-level attributes and stratified by baseline glycemic control: near normal (< 6.5%), good (6.5 - 8.0%), moderate (8.0 - 9.0%), and poor (>9.0%).

Results:

Although differences were not clinically significant, relative to no change in supermarket presence, supermarket loss was associated with worse A1c trajectories for those with good, moderate and poor baseline glycemic control. Supermarket gain was associated with marginally better A1c outcomes only among patients with near normal (<6.5%) A1c at baseline. In general, patients with the poorest glycemic control at baseline (>9.0%) had the worst associated changes in glycemic control following either supermarket loss or supermarket gain.

Conclusions:

For patients with diabetes, gaining neighborhood supermarket presence did not benefit glycemic control in a substantive way. The net health impacts of supermarket development and closure likely depend on a complex interaction of resident, neighborhood and store characteristics.

INTRODUCTION

The prevalence of type 2 diabetes mellitus is increasing and a major public health concern.[4,104] This chronic condition is characterized by insulin resistance and elevated blood sugar (hyperglycemia). Prolonged hyperglycemia can lead to crippling complications, including heart disease, stroke, blindness, kidney failure, and loss of digits or limbs. Eating a balanced diet rich in fruits, vegetables, whole grains, and lean proteins; and limited in sugar and refined carbohydrates can help improve glycemic control and prevent or delay these complications.[105]

The food retail options available where patients live may shape daily dietary choices and support or hinder individual efforts to self-manage this chronic condition.[30] Supermarkets generally offer a greater variety of healthy items, such as fresh produce, at lower prices relative to smaller grocery stores.[29] A growing number of local jurisdictions and states are funding policy initiatives to subsidize the development of supermarkets and other fresh food retail in underserved neighborhoods.[53,55,92,93] At the national level, the primary policy levers are the Healthy Food Financing Initiative,[52] a program that provides grants and subsidized loans for food retail investments, and the New Market Tax Credit, a program that provides tax credits for eligible businesses in low income areas.[106]

Many advocates of these policies implicitly assume that the introduction of supermarket retail will deliver positive health benefits for area residents; however this assumption lacks an evidence basis. Research on the health effects of community food retail interventions is still in nascent stages with few longitudinal analyses to date. In recent years, a handful of studies examined diet and weight outcomes before and after neighborhood changes in supermarket retail, but found no associations.[56,107–109]

Isolating the impact of the local food environment on health is challenging given that many factors influence the selection of residents and retail into neighborhoods, each of which may independently affect health. First, residential selection processes may bias study findings if people choose residential neighborhoods based on factors that are correlated with the outcome under investigation (e.g., individuals with healthier lifestyles may prioritize convenient access to supermarkets).[44] Retail selection is also problematic, because stores locate in neighborhoods based on factors correlated with individual health outcomes, e.g., local economic growth or population density.

Rigorous and thoughtful analyses are needed to help policy makers evaluate the merits of supermarket incentive programs and understand the health impact of neighborhood food retail changes on residents. Virtually nothing is known about how neighborhood food retail changes affect disease management outcomes among the most clinically vulnerable residents who already have a chronic condition. Given that eating a nutritious diet is critical for glycemic control, the opening of the first supermarket in a food desert or the closure of the last remaining neighborhood supermarket could have a major glycemic impact on diabetic residents. Moreover, among residents with diabetes, those with worse glycemic control may be differentially impacted by supermarket development and closure because they potentially have the most to gain from new access to fresh food retail and the most to lose from a loss of local fresh food retail.

Using data from the Kaiser Permanente Northern California Diabetes Registry over four years (2007-2010), we examined the impact of a gain or loss in the presence of a neighborhood supermarket on changes in glycemic control (based on hemoglobin A1c (A1c)) among patients with type 2 diabetes, stratified by baseline A1c levels. Relative to no change in neighborhood supermarket presence, we hypothesized that the opening of the first supermarket in a

neighborhood (hereafter referred to as "supermarket gain") would be associated with better A1c trajectories (i.e. larger A1c reductions or smaller A1c gains) and the closure of the last supermarket in a neighborhood (hereafter referred to as "supermarket loss") would be associated with worse A1c trajectories (i.e. smaller A1c reductions or greater A1c gains) in all strata. Moreover, we expected patients with worse A1c levels at baseline to have the greatest relative A1c reduction from supermarket gain and the greatest relative A1c increase from supermarket loss.

METHODS

STUDY SAMPLE

The sample was drawn from the Kaiser Permanente of Northern California (KPNC) Diabetes Registry[110], a dynamic cohort of KPNC health plan members with a history of diabetes mellitus ascertained from clinical and laboratory-based diagnoses. KPNC is a large, integrated not-for-profit health care delivery system that provides care for approximately one third of Northern California. A unique medical record number assigned to each member linked comprehensive clinical records with geospatial measures at the members' geocoded address at the Census block centroid of record. The study period spanned four years from Jan 1, 2007 to Dec 31, 2010 with annual measurements in each calendar year.

INCLUSION/EXCLUSION CRITERIA

Of all Registry members with type 2 diabetes, no end-stage renal disease, and at least one A1c record (n = 229,778 subjects), we excluded members with invalid address data (n = 11,579 subjects) as well as members with evidence of bariatric surgery, positive pregnancy test, end-stage renal disease or cancer diagnosis within one year of the study period (n = 33,676 subjects). Additionally, we restricted the sample of observations to limit residential mobility by retaining observations that correspond to years at each subject's longest address of residence (excluded n = 76,204 person-years). Thus, if a subject moved in the fourth year of the study period, we retained the first three years of observations. Alternatively, if a subject moved in the first year, we retained the last three years of observations. Finally, we excluded potential outliers by dropping observations with extreme annual A1c change in the top or bottom 0.5% of the frequency distribution. This left a final analytic sample of 160,000 subjects who contributed a total of 434,806 person-years.

OUTCOME: CHANGE IN A1C

Individual-level change in annual average hemoglobin A1c was the primary outcome of this study. A1c assays were ordered over the natural course of patient medical care and were conducted using high-performance liquid chromatography by a single KPNC central laboratory. For each subject, baseline A1c was defined as the annual, average A1c in the subject's first year of data in the study period .

EXPOSURE: CHANGE IN SUPERMARKET NEIGHBORHOOD PRESENCE

In each year of the study period, we mapped updated retail locations of supermarkets and large grocery stores (>$2 million in sales annually and >2,499 sq ft in floor area) based on InfoUSA business establishment data as distributed through ESRI Inc.[65] Store type designations within the commercial list were cleaned and reclassified based on Standard Industrial Codes (SIC) of businesses as well as by key word searches and name recognition. We created an indicator for supermarket presence in a neighborhood, defined as within a one-mile street network buffer around the census block centroid of each member's residential address, using ArcGIS 10.1.[66] Change in neighborhood presence of supermarket compared to the previous year was then transformed into three mutually exclusive indicator variables: supermarket gain, no change, and supermarket loss.

Supermarket gain refers to the development of the first supermarket in a neighborhood without any supermarkets in the previous year. Similarly, supermarket loss refers to the closure of the last neighborhood supermarket such that no neighborhood supermarkets are open in the current year. No change refers to the absence of both of these conditions. We developed these definitions under the assumption that with respect to neighborhood health, the marginal benefit provided by each new supermarket is greatest with the first neighborhood supermarket and the marginal health cost of each neighborhood supermarket closure is greatest with the last supermarket closure.

COVARIATES: CHANGE IN INDIVIDUAL AND NEIGHBORHOOD CHARACTERISTICS

Physical activity (PA) kernel density (e.g. parks, camps, gyms, and recreation centers) and unhealthful food outlet kernel density (fast food outlets and convenience stores) within a mile of each member's census block centroid were estimated using a quadratic function for inverse distance weighting. Additionally, Census block group measures of population density, poverty rate, and median self-reported home value were obtained from the 2005-2009, 2006-2010, 2007-2011, and 2008-2012 five-year aggregate American Community Survey (ACS) releases. Aggregate estimates were assigned to the middle year of the five-year period in panel data.[101]

The Charlson Comorbidity Index was calculated from prior year clinical records and serves as a validated measure of ten-year mortality risk based on the presence of 22 comorbid conditions.[99,100] Additionally, we created an indicator for the use of diabetes treatment medications including insulin, sulfonylurea, metformin and thiazolidinedione.

MISSING DATA

Missing values for both A1c and Charlson Comorbidity Index were imputed using chained equations as described in Appendix, Table 4. Given that standard errors were estimated using bootstrapping, a single stochastic imputation was drawn for each missing value in each bootstrap iteration.[70]

STATISTICAL ANALYSES

We estimated the effect of neighborhood supermarket presence change on individual-level A1c using a first-difference regression model[102] (Model 1) in STATA 12 (College Station, TX).[73]

Model 1:

$$A1c_{it} - A1c_{it-1} = \beta_0 + (SupermarketGain)\beta_1 + (SupermarketLoss)\beta_2 + (w_{it} - w_{it-1})\beta_3 + \varepsilon_{ij}$$

In which *i* subscripts the individual, *t* subscripts the year and w_{it} is a vector of covariates including Charlson Comorbidity Index, diabetes medication use, PA density, unhealthful food outlet density, population density, poverty rate and median census block group housing value. Quadratic and cubic polynomials were included for all continuous covariates to allow more flexible functional form. β_1 and β_2 estimate the effect of supermarket gain and supermarket loss on A1c change, respectively compared to the category of no presence change (the reference group).

Subjects were stratified by baseline A1c (< 6.5%, 6.5 - 8.0%, 8.0-9.0%, and ≥ 9.0%) and first difference models were estimated for each stratum separately. Since all observations for each subject reflect time at a single address (the address of longest residence), within-subject changes in supermarket exposure from year to year can be attributed to supermarket developments and closures rather than to changes in residence.

Observations for each subject were inverse-weighted by the subject's total contributed person-years of differenced observations to up-weight subjects with fewer years of data. To account for a correlated error structure, empirical confidence intervals were calculated from 1,000 nonparametric bootstrap iterations of the data imputation and model fitting process. Sensitivity analyses were conducted to test the robustness of estimates to more conservative model specifications that account for several potential sources of bias (details in Appendix, Table 6).

RESULTS

There was no significant change in A1c over time, although comorbidity scores did increase over the course of the study. Additionally median home prices declined, and poverty rate and population density increased over time. Changes in neighborhood supermarket presence were rare. Over the four year period, six percent of subjects experienced a gain in neighborhood supermarket where there were no supermarkets in the previous year and seven percent of subjects experienced a loss of neighborhood supermarket such that no supermarkets remained. Individual and neighborhood-level characteristics by baseline A1c strata are presented in Appendix Table 5.

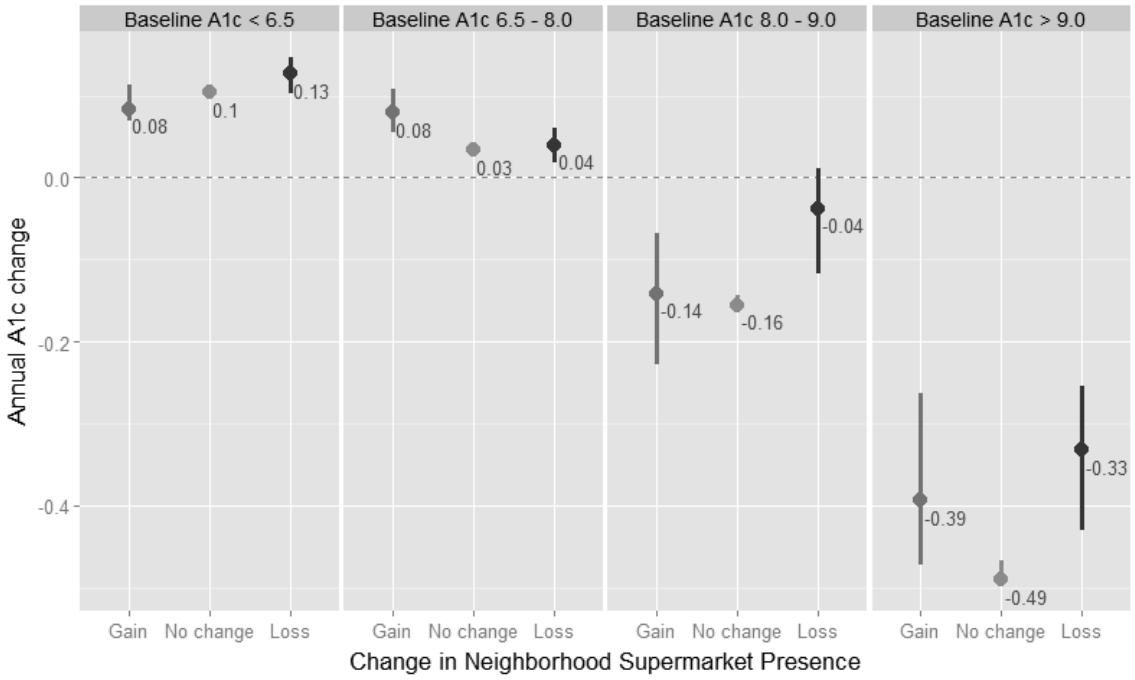
TABLE 1. INDIVIDUAL- AND NEIGHBORHOOD-LEVEL CHARACTERISTICS OF STUDY SAMPLE (POST-IMPUTATION)

	2007 (n = 143,152)	2008 (n = 143,152)	2009 (n = 146,324)	2010 (n = 145,328)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Individual-level variables				
Age	61.9 (12.6)	62.9 (12.6)	63.7 (12.6)	64.7 (12.6)
% Female	47.0%	46.8%	46.9%	46.8%
Race				
Asian	18.0%	18.0%	18.1%	18.0%
Black	9.3%	46.2%	9.4%	9.3%
Hispanic (Non-White)	20.0%	20.0%	19.9%	20.0%
White	46.2%	46.2%	46.0%	46.1%
Other	6.5%	6.5%	6.5%	6.5%
BMI	31.6 (7.1)	31.5 (7.2)	31.5 (7.2)	31.4 (7.2)
HbA1c	7.3 (1.4)	7.1 (1.4)	7.2 (1.4)	7.2 (1.4)
Comorbidity Score	1.8 (1.2)	2.0 (1.5)	2.2 (1.6)	2.3 (1.6)
% on insulin, TZDs, metformin, or sulfonylureas	66.0%	67.9%	67.4%	66.7%
Neighborhood-level variables				
% with any supermarket presence	54.0%	54.0%	52.6%	52.2%
% gained supermarket presence (vs prior year)	N/A	2.0%	2.5%	2.5%
% lost supermarket presence (vs prior year)	N/A	2.0%	3.8%	2.7%
Fast food & convenience store kernel density	1.2 (1.2)	1.1 (1.2)	1.2 (1.2)	1.2 (1.2)
Physical activity venue density	1.5 (1.7)	1.4 (1.7)	1.4 (1.7)	1.5 (1.7)
Median home price	$509,305	$483,864	$446,250	$413,092

	($216,743)	($227,225)	($232,101)	($235,830)
Poverty rate	10.0% (9.7%)	10.4% (10.3%)	11.0% (10.6%)	11.7% (10.9%)
Population density	8,084 (8,396)	8,298 (8,801)	8,354 (8,866)	8,396 (8,890)

Overall, patients with the lowest A1c at baseline experienced the greatest annual increase in A1c and patients with the highest A1c at baseline experienced the greatest annual decrease in A1c (Figure 1).

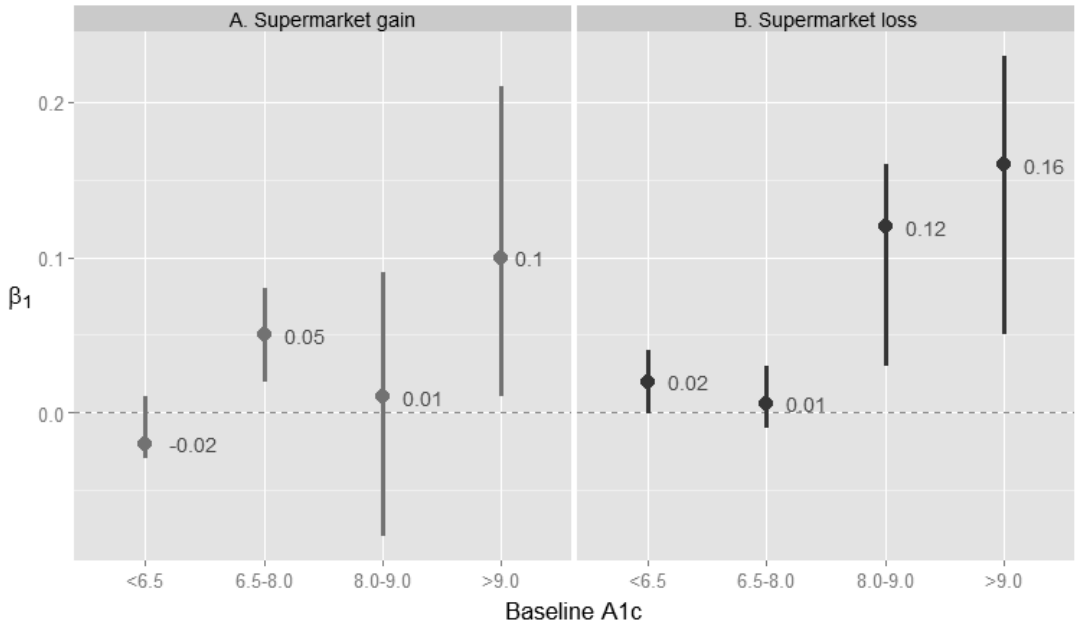
FIGURE 1. ANNUAL ADJUSTED A1C CHANGE BY BASELINE A1C AND NEIGHBORHOOD SUPERMARKET PRESENCE CHANGE



However, within each strata of baseline A1c, neighborhood supermarket presence change was associated with different magnitudes and directions of annual A1c change. Among patients with the lowest A1c, those with no change in supermarket presence experienced an average annual increase in A1c of 0.10%, 95% C.I. (0.10, 0.11), while those who gained neighborhood supermarket presence increased less [0.08% 95% C.I. (0.07, 0.11)] and those who lost neighborhood supermarket presence increased more [0.13%, 95% CI: (0.10, 0.15)]. This pattern shifts when we look at patients in other strata. Among patients with the highest A1c at baseline, those with no change in supermarket presence experienced a decrease in A1c of -0.49%, 95% C.I. (-0.47, -0.50), while those who gained or lost neighborhood supermarket presence decreased less; [-0.39%, 95% C.I. (-0.26, -0.47)] and [-0.33%, 95% C.I. (-0.25, -0.43)], respectively.

In strata-specific adjusted associations between neighborhood supermarket presence change, i.e. supermarket gain (β_1) and supermarket loss (β_2), and A1c change, the reference category is no presence change (Figure 2).

FIGURE **2**. ADJUSTED ASSOCIATIONS BETWEEN NEIGHBORHOOD SUPERMARKET PRESENCE CHANGE AND A1C CHANGE [A,B,C]



[A] Baseline A1c strata sample sizes: ≥ 9.0, n = 19,473 subjects; 8.0 - 9.0, n = 19,127 subjects; 6.5 - 8.0, n = 79,381 subjects; < 6.5, n = 51,590 subjects
[B] Adjusting for Charlson Comorbidity Index, diabetes medication use, PA density, unhealthful food outlet density, population density, poverty rate and median census block group housing value
[C] Percentile-based confidence intervals were obtained from 1,000 bootstrapped iterations of the data imputation and model estimation procedure

Supermarket gain was marginally associated with better A1c trajectories (smaller gains) relative to the reference category only among patients with near normal (<6.5%) glycemic control at baseline [-0.02%, 95% C.I. (-0.03, 0.01)]. For patients with good (6.5-8.0%) and poor (≥9.0%) A1c control, supermarket gain was associated with worse A1c trajectories (greater gains or smaller reductions) compared to the reference category [good: 0.05%, 95% C.I (0.02, 0.08); poor: 0.10%, 95% C.I. (0.01, 0.21)]. Compared to the reference category of no presence change, neighborhood supermarket loss was associated with worse A1c trajectories (greater gains or smaller reductions) for all patient strata. However, the magnitude of the association between supermarket loss and A1c change was greatest among patients with moderate [0.12%, 95% C.I. (0.03, 0.16] or poor glycemic control at baseline [0.16, 95% C.I. (0.05, 0.23)].

In sensitivity analyses (Appendix Table 6), we found similar results in unadjusted models, models adjusted for year effects and subject fixed effects, and in the complete case analysis. Additionally, the estimated associations between supermarket gain or loss and A1c change did not differ by neighborhood poverty rate. However, in the long-difference model which examined changes over four years, no significant associations were found.

DISCUSSION

This is the first longitudinal study to estimate the influence of supermarket availability on disease management outcomes for residents with a chronic condition. The sheer size of the KPNC analysis sample provided us with the power to detect small effect sizes that would have been statistically ambiguous in smaller studies. Furthermore, several of the measures and analytical methods employed in this study are relatively novel in the neighborhood health effects literature and may serve as a model for future research.

As we expected, relative to no change in supermarket presence, supermarket loss was associated with worse A1c trajectories. Also in line with expectations, patients with the poorest glycemic control at baseline appeared to have the largest relative A1c increase following supermarket loss. However, in contrast to our hypotheses, supermarket gain was associated with marginally better A1c trajectories only among patients with near normal (<6.5%) glycemic control at baseline. For patients in other strata, supermarket gain was associated with relative increases in A1c. In particular, patients with the poorest glycemic control at baseline experienced the most detrimental changes in A1c following a gain in neighborhood supermarket presence.

Although estimates were attenuated, these patterns resurfaced in alternate model specifications that adjusted for several potential sources of bias (Appendix Table 6). The associations that we found between neighborhood change in supermarket availability and A1c change were not explained by secular time trends or patient-level differences in A1c trajectories. Nor did the findings appear to be an artifact of the data imputation process. Moreover, while baseline A1c was strongly correlated with neighborhood poverty, the pattern of effect modification by baseline A1c does not appear to be attributable to underlying group differences in neighborhood affluence.

Overall, the estimated effects of supermarket presence gain or loss on A1c change were small relative to the effects of standard pharmaceutical and lifestyle recommendations such as metformin[111], exercise[112], and nutrition therapy[113] and would not be considered clinically significant. Additionally, it is important to note that the associations between supermarket gain or loss and A1c change were not observed in long-difference regression models and appear to be short-lived.

We propose several possible explanations for our findings. First, both gains and losses of neighborhood supermarket availability constitute major disruptions to the local retail environment and to household routines of food acquisition, preparation and consumption. A change in the food retail environment may trigger food purchasing changes as residents adapt to different shopping opportunities. Observed differences in effect estimates across baseline A1c strata may be due to differences in how patients adapt to this change. Patients with better control of their chronic condition may be better able to leverage the healthful food resources in the supermarket to improve their metabolic outcomes. On the other hand, patients who struggled to maintain glycemic control at baseline may have a harder time adjusting to either gains or losses in neighborhood supermarket availability.

Concurrently, observed differences in effect estimates across patient strata may be due to underlying differences in neighborhood supermarket attributes. Supermarkets vary widely by category (e.g. chain vs independent, ethnic vs non-ethnic, discount vs premium) and in-store characteristics (e.g. food availability, variety, product mix, and quality). Our study estimated an

average effect across all supermarkets and did not account for the influence of store-specific characteristics. This is a limitation of our analysis.

Lastly, our estimates may be confounded by unmeasured concurrent neighborhood changes. While we adjusted for changes in the density of fast food outlets and convenience stores, neighborhood supermarket gain may be accompanied by the development of other food retail, such as full-service restaurants and bars. These specific food retail changes were omitted in our study but may nonetheless affect A1c. Alternatively, supermarket closure may be a symptom of general neighborhood disinvestment leading to the deterioration or loss of unmeasured health assets such as a food bank or clinic. While we adjusted for changes in several neighborhood-level attributes, both inaccurate measurement of these covariates and omission of other time-varying neighborhood factors could bias our results.

Our study has several other limitations besides those already mentioned. First, this sample represents an insured population of type 2 diabetes patients who regularly obtain A1c assessments and results may not generalize to other populations. Although efforts were taken to minimize selection bias through missing data imputation and regression weighting, subjects who failed to contribute a single A1c assessment and subjects with non-geocodable addresses were excluded from analyses. Second, the time lag between neighborhood supermarket change and measurement of A1c is also an important consideration for the interpretation of our findings. Our annual measures of neighborhood retail reflect store changes in the preceding calendar year. Thus, the actual time-interval between neighborhood supermarket presence change and patient A1c assessment could vary from as short as one day to as long as 24 months.

Despite these limitations, we believe that this policy-relevant study makes a valuable contribution to the literature. In several influential cross-sectional studies, decreased proximity and availability of supermarkets have been associated with lower consumption of fruits and vegetables, poorer adherence to dietary guidelines, and greater risk of obesity.[31–34,37] While associations are not indicative of a causal relationship, these findings have shaped the public perception of supermarkets as community health assets.[114,115] Conversely, neighborhoods that lack supermarket presence (i.e. food deserts) are increasingly targets for policy intervention.[116]

Supermarket availability is commonly used as a proxy for neighborhood healthful food availability, but this practice is overly simplistic and may mislead policymakers to prioritize supermarket development over other policy solutions. Future longitudinal studies in this research domain should adopt more nuanced measures of healthy food availability which take into account within-store characteristics such as affordability, quality, product mix, and variety.

While access to healthful food is a necessary condition for the successful management of chronic condition, this evidence suggests that supermarket presence is not sufficient. Supermarkets offer a wide array of both healthful and unhealthful foods and the net health impacts of supermarket development and closure likely depend on a complex interaction of individual behavioral, neighborhood and store characteristics.

CHAPTER FIVE

DISCUSSION, SIGNIFICANCE AND FUTURE STEPS

CONCLUSION

As a whole, the three studies of this dissertation challenge the commonly held assumptions that 1) increasing physical proximity to supermarkets will improve the health outcomes of community residents and 2) neighborhood disparities in obesity are caused in part by the unequal geographical distribution of "healthful" and "unhealthful" food vendors. In Chapter 2, we found that the most deprived neighborhoods also had the greatest density of both healthful and unhealthful food outlets and disparities in the area density of food vendors did not explain the gap in obesity rates between less deprived and more deprived neighborhoods. In Chapter 3, among residents in the Kaiser Permanente Diabetes Registry who lived within two miles of 12 new supermarket developments, we failed to detect a relative improvement in BMI when the distance to nearest supermarket was reduced. Finally, in Chapter 4, we found that the opening of the first neighborhood supermarket and the closure of the last remaining neighborhood supermarket were both associated with different health outcomes for different diabetes patient subpopulations. The opening of the first neighborhood supermarket was associated with relatively better disease trajectories for residents with near normal glycemic control and relatively worse disease trajectories for residents with poor glycemic control. The closure of the last remaining supermarket in the neighborhood was associated with worse disease trajectories for all diabetes patients but the magnitude of the association was greatest for those patients with poor glycemic control at baseline.

Research on the health effects of neighborhood food access is methodologically challenging and the three studies in this dissertation demonstrated how methods borrowed from other fields of research could be adapted to address some of these challenges. The mediation analysis described in Chapter 2 is the first study to evaluate the pathway(s) through which neighborhood deprivation affects weight outcomes. Additionally, it is the only study that decomposes natural direct and indirect effects into component pathways, accounting for sequential impacts on intermediate variables along the causal chain. The effect decomposition traced the influence of the exposure through the causal sequence to better understand how the total effect of exposure on the outcome is mediated through various nodes.

Chapter 3 exploited 12 previous supermarket openings in Northern California as a "natural experiment" to investigate the impact of increasing supermarket proximity on BMI among patients in a large ongoing diabetes research study. While other studies in this field have also leveraged supermarket developments as imperfect proxies for experimental studies, these studies have all taken a prospective approach that necessitated data collection before and after the store opening. This approach leaves little room for error in timing and can be derailed when an anticipated supermarket does not open on schedule. This weakness was apparent in the recent pilot evaluation of a Philadelphia supermarket by Cummins and colleagues when a three year delay in the construction of the supermarket resulted in a three year difference in the follow-up time between residents in the intervention versus the comparison community. In contrast, the study detailed in Chapter 3 demonstrated how retrospective clinical and patient records could be leveraged for a similar purpose.

Finally, Chapter 4 demonstrated how first differencing, an econometric model that is relatively rare in the neighborhood health effects literature, could be used to examine how gains

and losses in neighborhood supermarket presence affect glycemic control for residents with type 2 diabetes. Additionally, the study used dummy indicators for every category of supermarket presence change (gain, loss, no change) which allowed for asymmetry in the magnitude and direction of effect estimates for a gain in supermarket presence versus a loss in supermarket presence. The potential asymmetry of neighborhood health effects is an issue that has not been adequately considered in existing studies. This additional flexibility in the operationalization of the exposure proved to be important because, in comparison to no change in neighborhood supermarket presence, both gains and losses of neighborhood supermarket presence were associated with worse A1c trajectories among patients with good, moderate, and poor glycemic control at baseline.

LIMITATIONS OF RESEARCH

This research has several limitations as described in the respective chapters. In each of the three studies, subjects represent an insured managed care population of adults with type 2 diabetes in a largely urban area, and study results may not be generalizable to populations that are dissimilar to the study sample. The primary limitation of the mediation analysis in Chapter 2 is the cross-sectional design of the study, meaning that there was no way to establish time ordering of the exposure, mediators and outcome. In the analyses presented in Chapters 3 and 4, observations of individuals following residential relocation were removed. This limits the generalizability of findings to a larger population because the characteristics of people who move may differ from those who stay. As is common in observational research, all of the analyses employed parametric statistical models. Misspecification of these models, either by omission of confounding variables or by miss-specification of functional form would lead to bias. Although each of the models adjusts for a wide range of confounding factors, residual confounding may remain. Additionally, with the exception of the analysis described in Chapter 3, ground-truth verification of supermarket and produce store locations was not feasible, resulting in potential inaccuracies in the characterization of food vendor and supermarket retail resources. All studies relied heavily on GIS measures of vendor density and proximity, which are not, on its own, appropriate proxies for neighborhood food access.[27,84,85] Within-store product characteristics are important components of food accessibility that may influence individual food choice and were not measured.

Despite these limitations, this research contributes to our understanding of how neighborhood deprivation translates into specific and potentially modifiable neighborhood factors that alter the daily existence of residents. It examined whether differences in food retail between rich and poor neighborhoods contribute to neighborhood-level health disparities, as is commonly believed. It estimated the impact on dietary quality and glycemic control of changes in supermarket proximity and neighborhood presence. The findings can help policy makers evaluate, prioritize, and design policy initiatives for addressing neighborhood disparities in diabetes and other chronic disease outcomes.

REFERENCES

1. Centers for Disease Control and Prevention. *Diabetes Report Card 2012*. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services; 2012. http://www.cdc.gov/diabetes/pubs/pdf/diabetesreportcard.pdf.

2. Centers for Disease Control and Prevention. *National Diabetes Statistics Report, 2014*. Atlanta, GA: Centers for Disease Control and Prevention; 2014. http://www.cdc.gov/diabetes/pubs/statsreport14/national-diabetes-report-web.pdf.

3. Centers for Disease Control and Prevention. *National Diabetes Fact Sheet: General Information and National Estimates in the United States, 2007.* Atlanta, GA: Centers for Disease Control and Prevention; 2009. http://stacks.cdc.gov/view/cdc/5613.

4. American Diabetes Association. Economic Costs of Diabetes in the U.S. in 2012. *Diabetes Care*. March 2013:DC_122625. doi:10.2337/dc12-2625.

5. Lindström J, Peltonen M, Eriksson JG, et al. High-fibre, low-fat diet predicts long-term weight loss and decreased type 2 diabetes risk: the Finnish Diabetes Prevention Study. *Diabetologia*. 2006;49(5):912-920. doi:10.1007/s00125-006-0198-3.

6. Diabetes Prevention Program Research Group. Reduction in the incidence of Type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002;346(6):393-403. doi:10.1056/NEJMoa012512.

7. Schwartz S, Fabricatore AN, Diamond A. Weight reduction in diabetes. *Adv Exp Med Biol*. 2012;771:438-458.

8. Skyler JS. Effects of Glycemic Control on Diabetes Complications and on the Prevention of Diabetes. *Clin Diabetes*. 2004;22(4):162-166. doi:10.2337/diaclin.22.4.162.

9. Evert AB, Boucher JL, Cypress M, et al. Nutrition Therapy Recommendations for the Management of Adults With Diabetes. *Diabetes Care*. October 2013. doi:10.2337/dc13-2042.

10. Alex ADA 1701 NBS, ria, 1-800-Diabetes V 22311. Making Healthy Food Choices. *Am Diabetes Assoc*. http://www.diabetes.org/food-and-fitness/food/what-can-i-eat/making-healthy-food-choices/. Accessed July 23, 2014.

11. Fitzgibbon M, Gans KM, Evans WD, et al. Communicating Healthy Eating: Lessons Learned and Future Directions. *J Nutr Educ Behav*. 2007;39(2, Supplement):S63-S71. doi:10.1016/j.jneb.2006.08.017.

12. Brambila-Macias J, Shankar B, Capacci S, et al. Policy interventions to promote healthy eating: A review of what works, what does not, and what is promising. *Food Nutr Bull*. 2011;32(4):365-375.

13. McKinnon RA, Orleans CT, Kumanyika SK, et al. Considerations for an Obesity Policy Research Agenda. *Am J Prev Med*. 2009;36(4):351-357. doi:10.1016/j.amepre.2008.11.017.

14. Diez R, Borrell L, Haan M, Jackson S, Schultz R. Neighbourhood environments and mortality in an elderly cohort: results from the cardiovascular health study. *J Epidemiol Community Health*. 2004;58(11):917-923. doi:10.1136/jech.2003.019596.

15. Chen JT, Rehkopf DH, Waterman PD, et al. Mapping and measuring social disparities in premature mortality: the impact of census tract poverty within and across Boston neighborhoods, 1999-2001. *J Urban Health Bull N Y Acad Med*. 2006;83(6):1063-1084. doi:10.1007/s11524-006-9089-7.

16. Drewnowski A, D. Rehm C, Solet D. Disparities in obesity rates: Analysis by ZIP code area. *Soc Sci Med*. 2007;65(12):2458-2463. doi:10.1016/j.socscimed.2007.07.001.

17. Roux AVD, Kershaw K, Lisabeth L. Neighborhoods and cardiovascular risk: Beyond individual-level risk factors. *Curr Cardiovasc Risk Rep*. 2008;2(3):175-180. doi:10.1007/s12170-008-0033-0.

18. Dubowitz T, Heron M, Bird CE, et al. Neighborhood socioeconomic status and fruit and vegetable intake among whites, blacks, and Mexican Americans in the United States. *Am J Clin Nutr*. 2008;87(6):1883-1891.

19. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health Bull N Y Acad Med*. 2006;83(6):1041-1062. doi:10.1007/s11524-006-9094-x.

20. Moffet HH, Adler N, Schillinger D, et al. Cohort Profile: The Diabetes Study of Northern California (DISTANCE)--objectives and design of a survey follow-up study of social health disparities in a managed care population. *Int J Epidemiol*. 2009;38(1):38-47. doi:10.1093/ije/dyn040.

21. Laraia BA, Karter AJ, Warton EM, Schillinger D, Moffet HH, Adler N. Place matters: neighborhood deprivation and cardiometabolic risk factors in the Diabetes Study of Northern California (DISTANCE). *Soc Sci Med 1982*. 2012;74(7):1082-1090. doi:10.1016/j.socscimed.2011.11.036.

22. Diez Roux AV. Investigating neighborhood and area effects on health. *Am J Public Health*. 2001;91(11):1783-1789.

23. Cohen DA, Farley TA, Mason K. Why is poverty unhealthy? Social and physical mediators. *Soc Sci Med*. 2003;57(9):1631-1641. doi:10.1016/S0277-9536(03)00015-7.

24. Leslie E, Cerin E, Kremer P. Perceived neighborhood environment and park use as mediators of the effect of area socio-economic status on walking behaviors. *J Phys Act Health*. 2010;7(6):802-810.

25. Boone-Heinonen J, Gordon-Larsen P. Obesogenic Environments in Youth: Concepts and Methods from a Longitudinal National Sample. *Am J Prev Med*. 2012;42(5):e37-e46. doi:10.1016/j.amepre.2012.02.005.

26. Giang T, Karpyn A, Laurison HB, Hillier A, Perry RD. Closing the grocery gap in underserved communities: the creation of the Pennsylvania Fresh Food Financing Initiative. *J Public Health Manag Pract JPHMP*. 2008;14(3):272-279. doi:10.1097/01.PHH.0000316486.57512.bf.

27. Charreire H, Casey R, Salze P, et al. Measuring the food environment using geographical information systems: a methodological review. *Public Health Nutr*. 2010;13(11):1773-1785. doi:10.1017/S1368980010000753.

28. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: A systematic review. *Health Place*. 2012;18(5):1172-1187. doi:10.1016/j.healthplace.2012.05.006.

29. Chung C, Myers SL. Do the Poor Pay More for Food? An Analysis of Grocery Store Availability and Food Price Disparities. *J Consum Aff*. 1999;33(2):276-296. doi:10.1111/j.1745-6606.1999.tb00071.x.

30. Horowitz CR, Colson KA, Hebert PL, Lancaster K. Barriers to Buying Healthy Foods for People With Diabetes: Evidence of Environmental Disparities. *Am J Public Health*. 2004;94(9):1549-1554.

31. Morland K, Wing S, Diez Roux A. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *Am J Public Health*. 2002;92(11):1761-1767.

32. Rose D, Richards R. Food store access and household fruit and vegetable use among participants in the US Food Stamp Program. *Public Health Nutr*. 2004;7(8):1081-1088. doi:10.1079/PHN2004648.

33. Laraia BA, Siega-Riz AM, Kaufman JS, Jones SJ. Proximity of supermarkets is positively associated with diet quality index for pregnancy. *Prev Med*. 2004;39(5):869-875. doi:10.1016/j.ypmed.2004.03.018.

34. Morland K, Diez Roux AV, Wing S. Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. *Am J Prev Med*. 2006;30(4):333-339. doi:10.1016/j.amepre.2005.11.003.

35. Liu GC, Wilson JS, Qi R, Ying J. Green neighborhoods, food retail and childhood overweight: Differences by population density. *Am J Health Promot*. 2007;21(4 SUPPL.):317-325.

36. Lopez RP. Neighborhood risk factors for obesity. *Obesity*. 2007;15(8):2111-2119. doi:10.1038/oby.2007.251.

37. Moore LV, Diez Roux AV, Nettleton JA, Jacobs DR. Associations of the Local Food Environment with Diet Quality—A Comparison of Assessments based on Surveys and Geographic Information Systems. *Am J Epidemiol*. 2008;167(8):917-924. doi:10.1093/aje/kwm394.

38. Cummins S, Petticrew M, Higgins C, Findlay A, Sparks L. Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. *J Epidemiol Community Health*. 2005;59(12):1035-1040. doi:10.1136/jech.2004.029843.

39. Boone-Heinonen J, Gordon-Larsen P, Kiefe CI, Shikany JM, Lewis CE, Popkin BM. Fast Food Restaurants and Food Stores: Longitudinal Associations With Diet in Young to Middle-aged Adults: The CARDIA Study. *Arch Intern Med*. 2011;171(13):1162-1170. doi:10.1001/archinternmed.2011.283.

40. Lee H. The role of local food availability in explaining obesity risk among young school-aged children. *Soc Sci Med*. 2012;74(8):1193-1203. doi:10.1016/j.socscimed.2011.12.036.

41. Elbel B, Moran A, Dixon LB, et al. Assessment of a government-subsidized supermarket in a high-need area on household food availability and children's dietary intakes. *Public Health Nutr*. February 2015:1-10. doi:10.1017/S1368980015000282.

42. Cummins S, Flint E, Matthews SA. New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. *Health Aff Proj Hope*. 2014;33(2):283-291. doi:10.1377/hlthaff.2013.0512.

43. Kawachi I, Subramanian SV. Neighbourhood influences on health. *J Epidemiol Community Health*. 2007;61(1):3-4. doi:10.1136/jech.2005.045203.

44. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med 1982*. 2004;58(10):1929-1952. doi:10.1016/j.socscimed.2003.08.004.

45. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc Sci Med*. 2004;58(10):1953-1960. doi:10.1016/S0277-9536(03)00414-3.

46. Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? *Int J Behav Nutr Phys Act*. 2007;4(1):32. doi:10.1186/1479-5868-4-32.

47. Sharkey JR, Horel S. Neighborhood Socioeconomic Deprivation and Minority Composition Are Associated with Better Potential Spatial Access to the Ground-Truthed Food Environment in a Large Rural Area. *J Nutr*. 2008;138(3):620-627.

48. Powell LM, Slater S, Mirtcheva D, Bao Y, Chaloupka FJ. Food store availability and neighborhood characteristics in the United States. *Prev Med*. 2007;44(3):189-195. doi:10.1016/j.ypmed.2006.08.008.

49. *Food, Conservation and Energy Act of 2008.*; 2008. http://www.usda.gov/documents/Bill_6124.pdf.

50. Agricultural Marketing Service, USDA. Creating Access to Healthy, Affordable Food. https://apps.ams.usda.gov/fooddeserts/foodDeserts.aspx. Accessed July 24, 2014.

51. Healthy Communities. *Lets Move*. http://www.letsmove.gov/healthy-communities. Accessed July 24, 2014.

52. Administration for Children and Families. CED Data Healthy Food Financing Initiative. *Off Community Serv*. January 2011. http://www.acf.hhs.gov/programs/ocs/resource/healthy-food-financing-initiative-0. Accessed July 23, 2014.

53. The Reinvestment Fund. The Pennsylvania Fresh Food Financing Initiative. http://www.thefoodtrust.org/php/programs/fffi.php. Accessed October 3, 2011.

54. The Colorado Health Foundation. Foundation in action: Colorado Fresh Food Financing Fund. 2014 2006. http://www.coloradohealth.org/CO4F/. Accessed July 23, 2014.

55. City of New Orleans. Fresh Food Retailer Initiative. *City New Orleans Mayor Mitchell J Landrieu*. October 2013. http://www.nola.gov/city/fresh-food-retailers-initiative/. Accessed July 23, 2014.

56. Boone-Heinonen J, Gordon-Larsen P, Kiefe CI, Shikany JM, Lewis CE, Popkin BM. Fast Food Restaurants and Food Stores: Longitudinal Associations With Diet in Young to Middle-aged Adults: The CARDIA Study. *Arch Intern Med*. 2011;171(13):1162-1170. doi:10.1001/archinternmed.2011.283.

57. Zhang YT, Laraia BA, Mujahid MS, et al. Does food vendor density mediate the association between neighborhood deprivation and BMI?: a G-computation mediation analysis. *Epidemiol Camb Mass*. 2015;26(3):344-352. doi:10.1097/EDE.0000000000000271.

58. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health*. 2001;55(2):111-122. doi:10.1136/jech.55.2.111.

59. Gustafson AA, Sharkey J, Samuel-Hodge CD, et al. Perceived and objective measures of the food store environment and the association with weight and diet among low-income women in North Carolina. *Public Health Nutr*. 2011;14(06):1032-1038. doi:10.1017/S1368980011000115.

60. Beydoun MA, Powell LM, Chen X, Wang Y. Food Prices Are Associated with Dietary Quality, Fast Food Consumption, and Body Mass Index among U.S. Children and Adolescents. *J Nutr*. 2011;141(2):304-311. doi:10.3945/jn.110.132613.

61. Jones-Smith JC, Karter AJ, Warton EM, et al. Obesity and the food environment: income and ethnicity differences among people with diabetes: the Diabetes Study of Northern California (DISTANCE). *Diabetes Care*. 2013;36(9):2697-2705. doi:10.2337/dc12-2190.

62. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7(9–12):1393-1512. doi:10.1016/0270-0255(86)90088-6.

63. Snowden JM, Rose S, Mortimer KM. Implementation of G-Computation on a Simulated Data Set: Demonstration of a Causal Inference Technique. *Am J Epidemiol*. 2011;173(7):731-738. doi:10.1093/aje/kwq472.

64. Gordon NP, Kaplan GA. Some evidence refuting the HMO "favorable selection" hypothesis: the case of Kaiser Permanente. *Adv Health Econ Health Serv Res*. 1991;12:19-39.

65. ESRI. http://www.esri.com/. Accessed February 24, 2014.

66. *ArcGIS Desktop*. Redland, CA: Environmental Systems Research Institute; 2012.

67. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381-1395. doi:10.1249/01.MSS.0000078924.61453.FB.

68. United States Census Bureau. 2000 Census. January 2012. http://www.census.gov/prod/www/decennial.html. Accessed February 24, 2014.

69. Peyrot M, Rubin RR. Structure and correlates of diabetes-specific locus of control. *Diabetes Care*. 1994;17(9):994-1001.

70. Daniel RM, Stavola BLD, Cousens SN. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *Stata J*. 2011;11(4):479-517.

71. Robins JM, Hernan MA. *Estimation of the Causal Effects of Time-Varying Exposures*. Vol (Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds.). Boca Raton: Crc Press-Taylor & Francis Group; 2009.

72. Daniel R m., Cousens S n., De Stavola B l., Kenward MG, Sterne J a. C. Methods for dealing with time-dependent confounding. *Stat Med*. 2013;32(9):1584-1618. doi:10.1002/sim.5686.

73. StataCorp. *Stata Statistical Software: Release 12.* College Station, TX: StataCorp LP; 2011.

74. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiol Camb Mass*. 2006;17(3):276-284. doi:10.1097/01.ede.0000208475.99429.2d.

75. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiol Camb Mass*. 2009;20(1):18-26. doi:10.1097/EDE.0b013e31818f69ce.

76. Fagerland M, Hosmer D. A generalized Hosmer-Lemeshow goodness-of-fit test for multinomial logistic regression models. *Stata J*. 2012;12(3):447-453.

77. StataCorp. Linktest. In: *Stata 12 Base Reference Manual*. Vol College Station, TX; 2011.

78. Zenk SN, Schulz AJ, Israel BA, James SA, Bao S, Wilson ML. Fruit and vegetable access differs by community racial composition and socioeconomic position in Detroit, Michigan. *Ethn Dis*. 2006;16(1):275-280.

79. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. *Am J Public Health*. 2006;96(2):325-331. doi:10.2105/AJPH.2004.058040.

80. Lepage B, Dedieu D, Savy N, Lang T. Estimating controlled direct effects in the presence of intermediate confounding of the mediator–outcome relationship: Comparison of five different methods. *Stat Methods Med Res*. October 2012:0962280212461194. doi:10.1177/0962280212461194.

81. Hoehner CM, Schootman M. Concordance of commercial data sources for neighborhood-effects studies. *J Urban Health Bull N Y Acad Med*. 2010;87(4):713-725. doi:10.1007/s11524-010-9458-0.

82. Auchincloss AH, Moore KAB, Moore LV, Diez Roux AV. Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. *Health Place*. 2012;18(6):1341-1347. doi:10.1016/j.healthplace.2012.06.016.

83. Powell LM, Han E, Zenk SN, et al. Field validation of secondary commercial data sources on the retail food outlet environment in the U.S. *Health Place*. 2011;17(5):1122-1131. doi:10.1016/j.healthplace.2011.05.010.

84. McKinnon RA, Reedy J, Morrissette MA, Lytle LA, Yaroch AL. Measures of the Food Environment: A Compilation of the Literature, 1990–2007. *Am J Prev Med*. 2009;36(4, Supplement):S124-S133. doi:10.1016/j.amepre.2009.01.012.

85. Caspi CE, Kawachi I, Subramanian SV, Adamkiewicz G, Sorensen G. The relationship between diet and perceived and objective access to supermarkets among low-income housing residents. *Soc Sci Med*. 2012;75(7):1254-1262. doi:10.1016/j.socscimed.2012.05.014.

86. VanderWeele TJ, Hernan MA. Causal inference under multiple versions of treatment. *J Causal Inference*. 2013;1(1):1-20. doi:10.1515/jci-2012-0002.

87. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiol Camb Mass*. 2011;22(3):368-377. doi:10.1097/EDE.0b013e3182109296.

88. Naimi AI, Kaufman JS, MacLehose RF. Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. *Int J Epidemiol*. May 2014:dyu107. doi:10.1093/ije/dyu107.

89. Kaufman JS. Commentary: Gilding the black box. *Int J Epidemiol*. 2009;38(3):845-847. doi:10.1093/ije/dyp163.

90. Schwartz S, Hafeman D, Campbell U, Gatto N. Author's Response: Gilding the Black Box. *Int J Epidemiol*. 2010;39(5):1399-1401. doi:10.1093/ije/dyp323.

91. California FreshWorks Fund. http://www.cafreshworks.com/. Accessed December 12, 2011.

92. Denver Food Access Task Force. *Healthy Food for All: Encouraging Grocery Investment in Colorado*. Denver, CO; 2011. http://www.coloradohealth.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=5625.

93. Office of Sustainability, Department of Planning. Baltimore Food Policy Initiative. 2010. http://archive.baltimorecity.gov/Government/AgenciesDepartments/Planning/BaltimoreFoodPolicyInitiative/FoodDeserts.aspx. Accessed April 2, 2015.

94. Infogroup.com. http://www.infogroup.com/. Accessed July 24, 2014.

95. Yelp. *Yelp*. http://www.yelp.com/. Accessed June 14, 2015.

96. Street View - Google Maps. https://www.google.com/maps/views/streetview?gl=us. Accessed June 14, 2015.

97. California ABC - License Query System. http://www.abc.ca.gov/datport/lqsmenu.html. Accessed June 14, 2015.

98. Moffet HH, Adler N, Schillinger D, et al. Cohort Profile: The Diabetes Study of Northern California (DISTANCE)--objectives and design of a survey follow-up study of social health disparities in a managed care population†. *Int J Epidemiol*. 2009;38(1):38-47. doi:10.1093/ije/dyn040.

99. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.

100. Van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. *J Clin Epidemiol*. 2010;63(7):798-803. doi:10.1016/j.jclinepi.2009.08.020.

101. US Census Bureau. About the American Community Survey. *U S Census Bur*. June 2014. http://www.census.gov/acs/www/about_the_survey/american_community_survey/. Accessed July 24, 2014.

102. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. Vol second edition edition. Cambridge, Mass: The MIT Press; 2010.

103. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2014. http://www.R-project.org.

104. Geiss LS, Wang J, Cheng YJ, et al. PRevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, united states, 1980-2012. *JAMA*. 2014;312(12):1218-1226. doi:10.1001/jama.2014.11494.

105. American Diabetes Association. Nutrition Recommendations and Interventions for Diabetes: A position statement of the American Diabetes Association. *Diabetes Care*. 2008;31(Supplement_1):S61-S78. doi:10.2337/dc08-S061.

106. The Reinvestment Fund. *New Market Tax Credits and Urban Supermarkets*.; 2011. http://www.cdfifund.gov/what_we_do/resources/NMTC%20for%20FUND%20approval%20101911.pdf.

107. Boone-Heinonen J, Diez-Roux AV, Goff DC, et al. The Neighborhood Energy Balance Equation: Does Neighborhood Food Retail Environment plus Physical Activity Environment = Obesity? The CARDIA Study. *Plos One*. 2013;8(12):e85141. doi:10.1371/journal.pone.0085141.

108. Cummins S, Flint E, Matthews SA. New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. *Health Aff Proj Hope*. 2014;33(2):283-291. doi:10.1377/hlthaff.2013.0512.

109. Elbel B, Moran A, Dixon LB, et al. Assessment of a government-subsidized supermarket in a high-need area on household food availability and children's dietary intakes. *Public Health Nutr*. 2015;FirstView:1-10. doi:10.1017/S1368980015000282.

110. Karter AJ, Schillinger D, Adams AS, et al. Elevated Rates of Diabetes in Pacific Islanders and Asian Subgroups The Diabetes Study of Northern California (DISTANCE). *Diabetes Care*. 2013;36(3):574-579. doi:10.2337/dc12-0722.

111. Hirst JA, Farmer AJ, Ali R, Roberts NW, Stevens RJ. Quantifying the effect of metformin treatment and dose on glycemic control. *Diabetes Care*. 2012;35(2):446-454. doi:10.2337/dc11-1465.

112. Umpierre D, Ribeiro PAB, Kramer CK, et al. Physical activity advice only or structured exercise training and association with HbA1c levels in type 2 diabetes: a systematic review and meta-analysis. *JAMA*. 2011;305(17):1790-1799. doi:10.1001/jama.2011.576.

113. Ajala O, English P, Pinkney J. Systematic review and meta-analysis of different dietary approaches to the management of type 2 diabetes. *Am J Clin Nutr*. 2013;97(3):505-516. doi:10.3945/ajcn.112.042457.

114. Treuhaft S, Karpyn A. *The Grocery Gap: Who Has Access to Healthy Food and Why It Matters*. Policy Link and The Food Trust; 2010. http://thefoodtrust.org/uploads/media_items/grocerygap.original.pdf. Accessed April 1, 2015.

115. Heinzmann D. Emanuel amplifies the call for more fresh food. *Chicago Tribune*. http://articles.chicagotribune.com/2011-10-25/news/ct-met-emanuel-food-desert-20111025_1_food-deserts-mari-gallagher-stores-in-underserved-areas. Published October 25, 2011. Accessed April 1, 2015.

116. United States, White House Task Force on Childhood Obesity, United States, Executive Office of the President. *Solving the Problem of Childhood Obesity within a Generation White House Task Force on Childhood Obesity Report to the President.* [Washington, D.C.]: Executive Office of the President of the United States; 2010. http://purl.access.gpo.gov/GPO/LPS122972. Accessed April 2, 2015.

APPENDIX

TABLE 1. MISSING DATA IMPUTATION

Imputation of missing data was performed with chained equations in STATA using the *mi impute chained* command with 10 burn-in cycles. Multiple linear regression was used to impute BMI values, which were inverse transformed, and the distributions of imputed and observed values were compared graphically. Logistic regression was used to impute dichotomous variables and multinomial logistic regression was used to impute categorical variables. All analytic variables were included as regressors in imputation models.

Variable	**Percent Missing**	**Imputation Method**
NDI	10.2%	Multinomial logistic regression
Healthy food vendor density	9.4%	Multinomial logistic regression
Unhealthy food vendor density	9.4%	Multinomial logistic regression
BMI (inverse-transformed)	4.2%	Linear multiple regression
Diet	17.3%	Logistic regression
Physical activity	15.7%	Logistic regression
Smoking	15.7%	Logistic regression
Race	3.9%	Multinomial logistic regression
Income to poverty ratio	15.5%	Multinomial logistic regression
Education	1.5%	Multinomial logistic regression
Household size	4.9%	Multinomial logistic regression
Marital status	4.3%	Multinomial logistic regression
Assets	22.0%	Logistic regression
Unemployment	3.2%	Logistic regression
Nativity	3.6%	Logistic regression
Internal locus of control	18.3%	Logistic regression
External locus of control	18.6%	Logistic regression
Population density	7.0%	Multinomial logistic regression

TABLE 2. PREDICTION MODELS

Variable	Model	Model Specification
Healthful vendor density (H_i)	Multinomial logistic	$ln\left(\frac{P(H_i=h)}{P(H_i=1)}\right) = \beta_0 + \beta_1 A_i + \beta_2 Ag_i + \beta_3 G_i + \beta_4 R_i + \beta_5 E_i + \beta_6 I_i + \beta_7 M_i + \beta_8 N_i + \beta_9 Em_i + \beta_{10} Hh_i + \beta_{11} L_i + \beta_{12} X_i + \beta_{13} V_i + \beta_{14} Pd_i + \beta_{15} A_i * Pd_i + \beta_{14} A_i * I_i + \beta_{15} A_i * R_i + \beta_{16} G_i * Ag_i + \beta_{17} Ag_i * E_i + \beta_{18} N_i * R_i + \beta_{19} N_i * Em_i + \beta_{20} E_i * Hh_i + \beta_{21} E_i * R_i + \beta_{22} A_i * E_i + \beta_{23} Ag_i * M_i + \beta_{24} S_i * I_i$
Unhealthful vendor density (U_i)	Multinomial logistic	$ln\left(\frac{P(U_i=u)}{P(U_i=1)}\right) = \beta_0 + \beta_1 A_i + \beta_2 Ag_i + \beta_3 G_i + \beta_4 R_i + \beta_5 E_i + \beta_6 I_i + \beta_7 M_i + \beta_8 N_i + \beta_9 Em_i + \beta_{10} Hh_i + \beta_{11} L_i + \beta_{12} X_i + \beta_{13} V_i + \beta_{14} Pd_i + \beta_{15} A_i * Pd_i + \beta_{16} A_i * I_i + \beta_{17} A_i * R_i + \beta_{18} Ag_i * E_i + \beta_{19} Ag_i * R_i + \beta_{20} N_i * R_i + \beta_{21} Em_i * Hh_i + \beta_{22} N_i * Em_i + \beta_{23} E_i * Hh_i + \beta_{24} E_i * R_i + \beta_{25} N_i * E_i + \beta_{26} A_i * E_i + \beta_{27} Ag_i * I_i + \beta_{28} A_i * Hh_i + \beta_{29} N_i * Ag_i + \beta_{30} G_i * I_i + \beta_{31} X_i * R_i$
Physically active (P_i)	Logistic	$ln\left(\frac{P(P_i=1)}{1-P(P_i=1)}\right) = \beta_0 + \beta_1 A_i + \beta_2 Ag_i + \beta_3 G_i + \beta_4 R_i + \beta_5 E_i + \beta_6 I_i + \beta_7 M_i + \beta_8 N_i + \beta_9 Em_i + \beta_{10} Hh_i + \beta_{11} L_i + \beta_{12} X_i + \beta_{13} V_i + \beta_{14} Pd_i + \beta_{15} A_i * Pd_i + \beta_{16} A_i * E_i + \beta_{17} Em_i * R_i + \beta_{18} Em_i * L_i + \beta_{19} E_i * R_i + \beta_{20} Ag_i * Hh_i + \beta_{21} N_i * Ag_i + \beta_{22} R_i * Hh_i$
Smoker (S_i)	Logistic	$ln\left(\frac{P(S_i=1)}{1-P(S_i=1)}\right) = \beta_0 + \beta_1 A_i + \beta_2 Ag_i + \beta_3 G_i + \beta_4 R_i + \beta_5 E_i + \beta_6 I_i + \beta_7 M_i + \beta_8 N_i + \beta_9 Em_i + \beta_{10} Hh_i + \beta_{11} L_i + \beta_{12} X_i + \beta_{13} V_i + \beta_{14} Pd_i + \beta_{15} A_i * I_i + \beta_{16} G_i * R_i + \beta_{17} Ag_i * R_i + \beta_{18} L_i * E_i + \beta_{19} N_i * E_i + \beta_{20} A_i * E_i + \beta_{21} Ag_i * Hh_i + \beta_{22} X_i * R_i + \beta_{23} R_i * Hh_i$
Diet adherence (D_i)	Logistic	$ln\left(\frac{P(D_i=1)}{1-P(D_i=1)}\right) = \beta_0 + \beta_1 A_i + \beta_2 H_i + \beta_3 U_i + \beta_4 P_i + \beta_5 S_i + \beta_6 Ag_i + \beta_7 G_i + \beta_8 R_i + \beta_9 E_i + \beta_{10} I_i + \beta_{11} M_i + \beta_{12} N_i + \beta_{13} Em_i + \beta_{14} Hh_i + \beta_{15} L_i + \beta_{16} X_i + \beta_{17} V_i + \beta_{18} Pd_i + \beta_{19} A_i * H_i + \beta_{20} A_i * R_i + \beta_{21} N_i * Em_i + \beta_{22} E_i * Q_i + \beta_{23} N_i * E_i + \beta_{24} A_i * E_i + \beta_{25} S_i * E_i + \beta_{26} P_i * Ag_i + \beta_{27} P_i * E_i$
BMI (Y_i^1)	Linear (inverse transformed)	$\frac{1}{Y_i^1} = \beta_0 + \beta_1 A_i + \beta_2 H_i + \beta_3 U_i + \beta_4 P_i + \beta_5 S_i + \beta_6 D_i + \beta_7 Ag_i + \beta_8 G_i + \beta_9 R_i + \beta_{10} E_i + \beta_{11} I_i + \beta_{12} M_i + \beta_{13} N_i + \beta_{14} Em_i + \beta_{15} Hh_i + \beta_{16} L_i + \beta_{17} X_i + \beta_{18} V_i + \beta_{19} Pd_i + \beta_{20} H_i * U_i + \beta_{21} G_i * R_i + \beta_{22} G_i * E_i + \beta_{23} Ag_i * E_i + \beta_{24} Ag_i * R_i + \beta_{25} N_i * R_i + \beta_{26} Em_i * R_i + \beta_{27} N_i * Em_i + \beta_{28} V_i * R_i + \beta_{29} A_i * D_i + \beta_{30} G_i * N_i + \beta_{31} Em_i * Hh_i + \beta_{32} Em_i * S_i + \beta_{33} E_i * R_i + \beta_{34} N_i * E_i + \beta_{35} P_i * R_i + \beta_{36} P_i * E_i + \beta_{37} S_i * R_i + \beta_{38} Ag_i * M_i + \beta_{39} D_i * G_i$
Obesity (Y_i^2)	Logistic	$ln\left(\frac{P(Y_i^2=1)}{1-P(Y_i^2=1)}\right) = \beta_0 + \beta_1 A_i + \beta_2 H_i + \beta_3 U_i + \beta_4 P_i + \beta_5 S_i + \beta_6 D_i + \beta_7 Ag_i + \beta_8 G_i + \beta_9 R_i + \beta_{10} E_i + \beta_{11} I_i + \beta_{12} M_i + \beta_{13} N_i + \beta_{14} Em_i + \beta_{15} Hh_i + \beta_{16} L_i + \beta_{17} X_i + \beta_{18} V_i + \beta_{19} Pd_i + \beta_{20} U_i * R_i + \beta_{21} Ag_i * R_i + \beta_{22} N_i * R_i + \beta_{23} Em_i * R_i + \beta_{24} V_i * R_i + \beta_{25} D_i * R_i + \beta_{26} P_i * S_i + \beta_{27} G_i * N_i + \beta_{28} Em_i * S_i + \beta_{29} E_i * R_i + \beta_{30} P_i * E_i + \beta_{31} S_i * R_i + \beta_{32} D_i * S_i + \beta_{33} Ag_i * M_i + \beta_{34} Ag_i * I_i + \beta_{35} L_i * Ag_i + \beta_{36} G_i * I_i + \beta_{37} M_i * R_i$
Severe obesity (Y_i^3)	Logistic	$ln\left(\frac{P(Y_i^3=1)}{1-P(Y_i^3=1)}\right) = \beta_0 + \beta_1 A_i + \beta_2 H_i + \beta_3 U_i + \beta_4 P_i + \beta_5 S_i + \beta_6 D_i + \beta_7 Ag_i + \beta_8 G_i + \beta_9 R_i + \beta_{10} E_i + \beta_{11} I_i + \beta_{12} M_i + \beta_{13} N_i + \beta_{14} Em_i + \beta_{15} Hh_i + \beta_{16} L_i + \beta_{17} X_i + \beta_{18} V_i + \beta_{19} Pd_i + \beta_{20} A_i * Pd_i + \beta_{21} G_i * R_i + \beta_{22} G_i * E_i + \beta_{23} Ag_i * R_i + \beta_{24} N_i * R_i + \beta_{25} Em_i * Hh_i + \beta_{26} N_i * Em_i + \beta_{27} D_i * R_i + \beta_{28} A_i * D_i + \beta_{29} Em_i * S_i + \beta_{30} E_i * R_i + \beta_{31} N_i * E_i + \beta_{32} P_i * E_i + \beta_{33} D_i * S_i + \beta_{34} Ag_i * I_i + \beta_{35} L_i * Ag_i + \beta_{36} N_i * Ag_i + \beta_{37} G_i * I_i + \beta_{38} X_i * R_i + \beta_{39} R_i * Hh_i + \beta_{40} E_i * Hh_i$

A_i: NDI, Ag_i: age, G_i: gender, R_i: race, Ed_i: education, I_i: income, V_i: value of assets, M_i: marital status, N_i: nativity, Em_i: employment status, Hh_i household size, L_i: internal locus of control, X_i: external locus of control, Pd_i: population density

TABLE 3. SUBCOMPONENT MEDIATION PARAMETER DEFINITIONS

Parameter	Definition
a	$E\{H(A=1)-H(A=4)\}$
b	$E\{U(A=1)-U(A=4)\}$
c	$E\{D(A=4,P(A=4),S(A=4),H(A=1),U(A=4)) - D(A=4,P(A=4),S(A=4),H(A=4),U(A=4))\}$
d	$E\{D(A=4,P(A=4),S(A=4),H(A=4),U(A=1)) - D(A=4,P(A=4),S(A=4),H(A=4),U(A=4))\}$
e	$E\{D(A=1,P(A=1),S(A=1),H(A=1),U(A=1)) - D(A=4,P(A=4),S(A=4),H(A=1),U(A=1))\}$
f	$E\{Y[A=4,P(A=4),S(A=4),H(A=1),U(A=4),D(A=4,P(A=4),S(A=4),H(A=4),U(A=4))] - Y[A=4,P(A=4),S(A=4),H(A=4),U(A=4),D(A=4,P(A=4),S(A=4),H(A=4),U(A=4))]\}$
g	$E\{Y[A=4,P(A=4),S(A=4),H(A=4),U(A=1),D(A=4,P(A=4),S(A=4),H(A=4),U(A=4))] - Y[A=4,P(A=4),S(A=4),H(A=4),U(A=4),D(A=4,P(A=4),S(A=4),H(A=4),U(A=4))]\}$
h	$E\{Y[A=4,P(A=4),S(A=4),H(A=4),U(A=4),D(A=4,P(A=4),S(A=4),H(A=1),U(A=1))] - Y[A=4,P(A=4),S(A=4),H(A=4),U(A=4),D(A=4,P(A=4),S(A=4),H(A=4),U(A=4))]\}$
i	$E\{Y[A=1,P(A=1),S(A=1),H(A=1),U(A=1),D(A=1,P(A=1),S(A=1),H(A=1),U(A=1))] - Y[A=4,P(A=4),S(A=4),H(A=1),U(A=1),D(A=1,P(A=1),S(A=1),H(A=1),U(A=1))]\}$
J	$E\{Y[A=1,P(A=1),S(A=1),H(A=1),U(A=1),D(A=1,P(A=1),S(A=1),H(A=1),U(A=1))] - Y[A=1,P(A=1),S(A=1),H(A=1),U(A=1),D(A=4,P(A=4),S(A=4),H(A=1),U(A=1))]\}$

TABLE 4. MISSING DATA IMPUTATION

Imputation of missing data was performed with chained equations in STATA using the *mi impute chained* command with 10 burn-in cycles. Multiple linear regression was used to impute HbA1c values (inverse-square transformed) and the Charlson comorbidity score. All model covariates and HbA1c and Charlson scores from other years were included as regressors in imputation models. Annual average BMI value and a kernel density measure for all businesses were included in imputation models as auxiliary variables. Imputed values were trimmed to be within the range of observed values and imputed Charlson scores were rounded to the nearest integer.

	% Missing			
Variable	**2007**	**2008**	**2009**	**2010**
HbA1c				
Pre-imputation	14.9%	16.2%	21.6%	24.4%
Post-imputation	3.4%	4.5%	6.5%	7.0%
Charlson comorbidity				
Pre-imputation	8.6%	7.0%	9.2%	15.5%
Post-imputation	2.5%	3.0%	2.8%	4.8%
On insulin, sulfonylureas, TZDs	0.0%	0.0%	0.0%	0.0%
One mile supermarket presence	0.0%	0.0%	0.0%	0.0%
One mile kernel density				
Physical activity	0.0%	0.0%	0.0%	0.0%
Convenience stores and fast food	0.0%	0.0%	0.0%	0.0%
ACS block group values				
Poverty rate	0.0%	0.0%	0.0%	0.0%
Median home value	0.0%	0.0%	0.0%	0.0%
Population density	0.0%	0.0%	0.0%	0.0%

TABLE 5. CHARACTERISTICS OF STUDY SAMPLE (IN 2007) BY BASELINE A1C STRATA

	A1c < 6.5% (n =43,521)	A1c 6.5-8.0% (n = 68,358)	A1c 8.0-9.0% (n = 16,098)	A1c ≥ 9.0% (n = 15,175)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Individual-level variables				
Age***	64.6 (12.7)	62.6 (12.1)	58.3 (11.6)	54.6 (11.4)
% Female***	46.3%	48.3%	44.7%	43.7%
Race				
Asian***	14.0%	20.6%	19.3%	16.4%
Black***	7.8%	9.2%	10.5%	12.8%
Hispanic***	16.6%	18.9%	24.0%	30.3%
White***	55.1%	44.6%	39.7%	34.4%
Other	6.4%	6.7%	6.5%	6.1%
BMI***	30.9 (6.8)	31.6 (7.1)	32.5 (7.2)	32.8 (7.4)
Comorbidity Score***	1.9 (1.3)	1.8 (1.2)	1.8 (1.2)	1.7 (1.1)
% on insulin, TZDs, metformin or sulfonylureas***	44.2%	72.8%	82.5%	80.8%
Neighborhood-level variables				
% with supermarket presence***	53.2%	54.2%	53.9%	55.3%
Fast food & convenience store kernel density***	1.1 (1.3)	1.2 (1.2)	1.2 (1.2)	1.2 (1.2)
Physical activity venue density*	1.4 (1.6)	1.5 (1.7)	1.5 (1.7)	1.5 (1.7)
Median home price***	$521,511 ($223,486)	$513,376 ($215,911)	$493,040 ($210,805)	$481,162 ($203,680)
Poverty rate***	9.5% (9.5%)	9.8% (9.6%)	10.6% (10.1%)	11.3% (10.5%)
Population density***	7,591 (7,863)	8,155 (8,517)	8,387 (8,792)	8,727 (8,711)

Stars indicate that differences across strata are significant at the p<0.05 level (*), p<0.01 level (**), or p<0.001 level (***).

TABLE 6. SENSITIVITY ANALYSES

Model	# subjects (observations)	Supermarket gain β_1	95% CI[a]	Supermarket loss β_2	95% CI[a]
Model 0: Unadjusted					
A1c ≥ 9.0	18,343 (47,125)	0.10	[0.01, 0.21]	0.18	[0.08, 0.27]
A1c 8.0 - 9.0	18,013 (48,693)	0.02	[-0.06, 0.09]	0.11	[0.03, 0.16]
A1c 6.5 - 8.0	75,123 (207,390)	0.05	[0.02, 0.08]	0.01	[-0.02, 0.03]
A1c < 6.5	48,521 (133,513)	-0.02	[-0.03, 0.01]	0.03	[0.00, 0.05]
Model 2: Model 1 + year effects					
A1c ≥ 9.0	18,343 (46,860)	0.07	[-0.01, 0.19]	0.08	[-0.01, 0.16]
A1c 8.0 - 9.0	18,013 (48,485)	0.01	[-0.08,0.08]	0.09	[0.01, 0.14]
A1c 6.5 - 8.0	75,123 (206,535)	0.04	[0.02, 0.07]	-0.01	[-0.03, 0.01]
A1c < 6.5	48,521 (132,926)	-0.02	[-0.04,0.01]	0.02	[-0.01, 0.04]
Model 3: Model1 + subject fixed effects					
A1c ≥ 9.0	18,343 (46,860)	0.11	[0.00, 0.25]	0.20	[0.06, 0.27]
A1c 8.0 - 9.0	18,013 (48,485)	0.06	[-0.07, 0.13]	0.11	[-0.02, 0.14]
A1c 6.5 - 8.0	75,123 (206,535)	0.04	[0.02, 0.09]	0.01	[-0.01, 0.04]
A1c < 6.5	48,521 (132,926)	-0.02	[-0.04, 0.01]	0.02	[-0.01, 0.05]

Model 4: Model 1 with complete case sample					
A1c ≥ 9.0	14,896 (30,374)	0.07	[-0.03, 0.23]	0.14	[0.01,0.25]
A1c 8.0 - 9.0	15,470 (34,529)	0.04	[-0.06, 0.13]	0.06	[-0.03,0.15]
A1c 6.5 - 8.0	66,821 (156,964)	0.07	[0.03,0.10]	0.01	[-0.02,0.03]
A1c < 6.5	42,341 (95,559)	-0.01	[-0.03, 0.02]	0.03	[0.00, 0.05]
Model 5: Model 1 by poverty rate strata					
Poverty ≥ 20%	23,272 (61,612)	0.03	[-0.02, 0.08]	0.05	[-0.02, 0.09]
Poverty < 20%	136,728 (373,194)	0.02	[0.01, 0.05]	0.04	[0.02, 0.06]
Model 6: Long difference (four-year)					
A1c ≥ 9.0	14,017 (14,017)	-0.06	[-0.13, 0.13]	-0.04	[-0.16, 0.08]
A1c 8.0 - 9.0	14,950 (14,950)	-0.04	[-0.11, 0.10]	0.01	[-0.06, 0.11]
A1c 6.5 - 8.0	64,045 (64,045)	0.03	[-0.03, 0.05]	0.02	[-0.02, 0.04]
A1c < 6.5	40,958 (40,958)	0.00	[-0.03, 0.03]	0.00	[-0.03, 0.03]

Model 0 : Unadjusted associations.
Model 2: Full model with year fixed effects to account for potential secular changes in A1c across time.
Model 3: Full model with individual fixed effects such that subjects serve as their own controls not only with respect to their annual A1c levels but also with respect to their A1c trajectories over time.
Model 4: Full model excluding observations with any missing values to examine the impact of the data imputation process on estimates.
Model 5: Full model stratified by neighborhood poverty rate to explore whether the differences in estimates across A1c strata are attributable to underlying differences in neighborhood poverty.
Model 6: Long difference model to estimate the effect of supermarket change over a longer time horizon. All model variables are defined as change over four years (2007-2010).