Investigating food environment disparities using digital supermarket transactions for fruits and vegetables in Quebec

Deepa Jahagirdar
Epidemiology, Biostatistics & Occupational Health
McGill University, Montreal
June 2018

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of PhD Epidemiology

Table of Contents

I Abstract and Résumé

Introduction

Consuming fruits and vegetables can help to prevent chronic disease and illness. Despite widespread awareness campaigns, socioeconomic disparities in fruit in vegetable consumption remain. To address this issue, many governments in North America and Europe have aimed to improve geographical access to stores selling healthy foods in vulnerable areas, mainly supermarkets. However, there is little to suggest that having access to more supermarkets improves diet. As a result, calls to shift away from accessibility and enhance research related to other food environment dimensions are increasingly prevalent. These dimensions include affordability (cost), availability (the selection of healthy items) and variety (the varieties of each available item).

In this manuscript-based thesis I used six years (2008-2013) of weekly automatically-collected digital grocery store transaction data to study the under-investigated dimensions of the food environment.

Aims

In the first aim, I determined whether areas with more supermarkets have more affordable fruits and vegetables. In the second aim, I determined whether the availability and variety of fruits and vegetables depended on chains' average price levels and regional income. In the third aim, I quantified seasonal and regional disparities and fruit and vegetable affordability.

Results

First, I found that the number of supermarkets in a region did not necessarily explain the affordability of fruits and vegetables; the chain make-up and degree of rurality were more important. Second, I found that the availability and variety of different fruit and vegetable products were lower at low price chains. Finally, I found that high price regions had greater seasonal fluctuations than low price regions.

Conclusion

Digital scanner data is an alternative to in-store measurement that could enhance future research on the food environment. In this thesis, I used these data to find substantial between-supermarket and seasonal variation in affordability, availability and variety of fruits and vegetables. Policies that focus mainly on access to 'healthy' food retailers may not succeed if they fail to simultaneously consider other food environment characteristics.

Introduction

Consommer des légumes et fruits au quotidien pourrait réduire le risque de plusieurs maladies chroniques. Malgré les campagnes de promotion de la santé, des disparités socio-économiques dans la consommation demeurent évidentes. Pour faire face au problème, de nombreux gouvernements en Amérique du Nord et en Europe ont fondé des interventions axées sur l'accessibilité, principalement à travers un meilleur accès géographique aux supermarchés. Toutefois, peu de données probantes démontrent une association de la présence de supermarchés à une consommation accrue de fruits et légumes. Par conséquent, il y a un besoin d'études portant sur d'autres dimensions de l'environnement alimentaire. Ces dimensions incluent le prix, les choix santé offerts et la variété des produits offerts. Dans cette thèse, j'utilise six ans (2008-2013) de données numériques des transactions de supermarchés au Québec pour étudier ces dimensions de l'environnement alimentaire.

Objectifs

Le premier objectif s'attarde à mesurer l'association entre l'accessibilité et les prix régionaux de fruits et légumes. Le deuxième objectif vise à déterminer si les supermarchés moins chers offrent moins de choix et de variété de fruits et légumes. Finalement, je quantifie la variation saisonnière et régionale dans le troisième objectif.

Résultats

Premièrement, les régions avec plus de supermarchés ne sont pas nécessairement moins chères. La direction de l'association dépend de la ruralité et des chaînes qui sont présentes. Deuxièmement, les commerces de faible prix offrent moins de choix et de variété de fruits et légumes. Finalement, les disparités saisonnières étaient importantes. Les régions plus chères montraient aussi une plus grande amplitude saisonnière que les autres régions.

Conclusions

Les données de transactions des supermarchés offrent une alternative aux études *in situ* pour mesurer l'environnement alimentaire. Dans cette thèse j'ai utilisé ces données pour démontrer qu'il existe une hétérogénéité substantielle de prix, de choix offerts et de variété de fruits et légumes entre les supermarchés, régions et saisons. En outre, les politiques qui se concentrent sur l'accessibilité pourraient être moins efficace si elles ne considèrent pas les autres dimensions de l'environnement alimentaire simultanément.

II Acknowledgements

My sincere thanks go to my two supervisors Drs David Buckeridge and Amélie Quesnel-Vallée. Their patience with the long process of data cleaning and structuring prior to any interesting analyses, and subsequent feedback on my ideas in this uncharted area of research, was invaluable. Further, their support for training opportunities and conferences ensured that I could place the work in the broader context of latest developments in this field. I also thank my committee member, Dr Yu Ma, for his unique insight related to marketing concepts and using marketing data.

I am also grateful for Drs Arijit Nandi and Sam Harper for providing the opportunity for a fellowship with the Institute for Health and Social Policy. The training in methods to analyse clustered data I received as part of work with them formed the basis for the methods I ended up using in this thesis. Dr James Hanley also provided helpful statistical advice on clustering for the 1st study.

My labmates in the Institute for Clinical and Health Informatics and my PhD cohort also provided countless hours of conversation related to methods, study design, PhDs and life. They made the past five years bearable and forced me to question almost every methodologic and analytic decision I made. Specifically, Aman Verma and Guido Powell's help with endless coding, logistical and programmatic issues related to learning the software and hardware required to analyse 'big data' was vital. Hiroshi Mamiya and Luc de Montigny's previous work on the business directory facilitated my work as well. Hiroshi's prior experience with marketing data also helped me to situate this thesis in existing literature and ensure the ideas were truly innovative.

Finally, thanks to my friends and family for their continued support for whatever I want to do. In particular, my husband Prasad's support staying up late via video chat many nights while I finished thesis sections or vented about the difficulty in staying on track. The PhD process was not as hard as might have been as a result.

III Statement of Originality

This PhD thesis makes fundamental methodologic and theoretical contributions to research on food environments. Despite the large quantity of research since 2005, progress has been limited by the lack of tools to measure the food environment and the use of simple methods where more advanced techniques are indicated. These gaps have prevented estimating the impact of the food environment on diets and health outcomes. In this thesis, I use automatically-collected digital grocery store scanner data to measure food environment attributes that were previously only measured using resource intensive in-store surveys. Further, I used more advanced analytic techniques to derive interpretable measures and ensure my results appropriately accounted for the temporal and spatial nature of these data.

In the first manuscript, I challenge a commonly held assumption in this literature that supermarkets inherently improve the regional affordability of fruits and vegetables. In the second manuscript, I measure availability and variety which have rarely ever been described, to demonstrate differences in the types of fruits and vegetables supermarkets with different average prices offer. In the third manuscript, I use a method previously used for influenza surveillance to demonstrate regional and seasonal variation in fruit and vegetable affordability, which has again not been done before. The methods and concepts used and addressed in this thesis lay groundwork for future research that could link food environment dimensions to health outcomes. They also provide a fresh perspective on conducting research in this field, which could lead to better evidence to inform policies that target the food environment.

IV Author Contributions

Manuscript 1

Jahagirdar, D; Quesnel-Vallée, A, Ma, Y, Buckeridge, D. (2018) The dynamic between affordability and accessibility of healthy food. *Submission planned for American Journal of Preventive Medicine.*

Deepa Jahagirdar (DJ) proposed the study, carried out the analyses, and wrote the paper. Amélie Quesnel-Vallée's (AQV) expertise in clustered data methods helped to inform the analysis. Yu Ma's (YM) familiarity with the marketing data helped to define the variables, and generate the market weights. David Buckeridge's (DB) input was mainly regarding the geographic boundaries, additional sensitivity analysis and modelling time. The co-authors all provided useful feedback.

Manuscript 2

Jahagirdar, D; Quesnel-Vallée, A, Ma, Y, Buckeridge, D.(2018) Disparities in the availability and variety of fruits and vegetables. *Submission planned for American Journal of Public Health.*

DJ proposed the study, carried out the analyses, and wrote the paper. YM's familiarity with the marketing concepts helped to ground availability and variety in pre-existing literature. AQV and DB provided useful feedback in terms of the structure and framing the paper to directly address a public health problem.

Manuscript 3

Jahagirdar, D; Quesnel-Vallée, A, Ma, Y, Buckeridge, D. (2018) Temporal and Seasonal Variation in the Affordability of Fruits and Vegetables. *Submission planned for International Journal of Behavioural Nutrition and Physical Activity.*

DJ proposed the study, carried out the analyses, and wrote the paper. DB's expertise in surveillance methods helped to finalize the methods. AQV and YM helped with the conceptual framing of the paper and all authors reviewed several drafts.

V Financial Support

My first year of study was funded by the Department of Epidemiology, Biostatistics, and Occupational Health at McGill University, and a summer internship at the *Institut National de Santé Publique du Québec* was funded by Dr Gilles Paradis. I subsequently received two *Max E Binz* fellowships from the Faculty of Medicine which funded my 2nd and 4th years, with matching funds from Dr Amélie Quesnel-Vallée and David Buckeridge, respectively. My 3rd and half the 4th year were funded through the MachEquity Doctoral Fellowship with Dr Arijit Nandi. I additionally received a GREAT award to present my work at the Informatics for Health conference in 2017 in Manchester, UK. David Buckeridge supplemented the rest of the funding for the conference, and also funded a training opportunity at the Institute for Big Data at the University of Washington.

VI List of Abbreviations

AIC = Akaike Information Criterion

ARIMA = Autoregressive Integrated Moving Average

BIC = Bayesian Information Criterion

CI = Confidence interval

FSA = Forward Sortation Area

mFREI = Modified Food Retail Environment Index

NEMS-S = Nutrition Environment Measures Survey-Stores

SD = Standard Deviation

SIC = Standard Industry Classification

RLS = *réseaux locaux de services*

VII List of Figures and Tables

Figures

Tables

Equations

Chapter 1 Introduction

In 2010, the Obama Administration committed to invest in bringing healthy food retailers like supermarkets to underserved areas with the goal of improving food environments.[1] Major cities in the U.S. have issued their own laws and policies to improve geographic access to healthy food.[2,3] In the UK, improving shopping access in deprived areas was part of a national strategy for neighbourhood renewal to reduce social exclusion.[5] Canada has followed suit. A Health Canada task force has established indicators to characterize geographic food access[6] and major Canadian cities have implemented their own policies to target low access to healthy food.[8] For example, Toronto has planned to invest in healthy corner stores that sell fresh fruits and vegetables.[8]

Governments in North America and Europe have prioritized improving the food environment because poor diet is accepted as a driver of obesity and chronic disease. In particular, people worldwide do not eat enough fruits and vegetables.[9] This food group is protective against a slew of chronic diseases and conditions including ischaemic heart disease, stroke, and cancers of the stomach, esophagus, lung, colon, and rectum.[10] Encouraging higher fruit and vegetable consumption has been the subject of countless government and industry campaigns.[11] While they have succeeded in increasing awareness, such campaigns have not increased consumption across all socio-economic classes.[12-14] The search for new approaches that could improve population diet and reduce dietary inequalities is occurring against the failure of awareness campaigns. It is part of a general trend towards understanding and addressing environmental causes of unhealthy lifestyles. The food environments literature has proliferated in this context.

Despite the policy enthusiasm, food environments is a research field in its infancy.[15] Glanz et al's[15] 2005 framework catalyzed the expansion of the evidence base. She divided the food environment into the *consumer* versus the *community nutrition environment* to distinguish between within-store factors such as price and variety of products versus structural factors such as the number of supermarkets in a given area. The number of policies and initiatives targeting store access reflects the disproportionate focus on the community nutrition environment in the literature. However, the evidence suggesting that this environment is worse in deprived areas or associated with health outcomes is weak,[16,17] particularly in Canada.[18] More evidence points to the importance of within-store factors including affordability to increase consumption,[19] which may also be a more important source of disparity. But these factors have not been given as much attention because standard methods for assessing within-store factors require resource-intensive visits to stores to survey prices and the products on offer.

My thesis aims to address conceptual and methodological limitations in this literature. Underlying all the papers, I make innovative use of weekly, automatically-collected digital grocery store transaction data. These data provide weekly records of prices, sales and products sold at a random sample of about 200 supermarkets in Quebec, Canada. The large quantity of high spatial and temporal resolution data helps to overcome the typical time and cost pressures associated with in-store surveys to measure within-store characteristics. I use these data to: i) determine the association between the affordability of fruits and vegetables and accessibility of supermarkets; ii) determine differences in fruit and vegetable product variety in low, medium and high price chains; and iii) measure the seasonal difference in fruit and vegetable affordability overall, and in different regions. In doing so, I hope to contribute to the evidence on food environment dimensions that have remained unexplored to date.

Chapter 2 Background

A report from the 1943 United Nations Conference on Food Agriculture provided a groundbreaking suggestion: food consumption at a level sufficient to prevent malnutrition is not enough to promote health and wellbeing.[12] This paradigm shift reached Canada in 1982. Dietary recommendations changed from concerns about ingesting enough to concerns about chronic disease.[20] The change reflected the earliest associations between excess consumption of fats and cholesterol, and coronary heart disease and cancer.[21] Today, over 35 million people die each year from chronic diseases such as heart disease, stroke and cancer.[22] The evidence has continued to escalate to demonstrate links between what we eat and several types of cancer,[23] poorer mental health[24] and cognitive performance.[25] Sugar has also received increased attention due to its association with several conditions,[26] including hypertension, non-alcoholic fatty liver disease and other signs of metabolic dysfunction[27] and potentially Alzheimer's disease.[28] Poor dietary habits are accepted as a major contributor to non-communicable diseases.

By the late 1980s fruit and vegetable consumption was linked to reduced cancer and other chronic disease risk.[29] This food group is low on the common dietary substances associated with disease and illness, and high on protective elements. Aside from containing low amounts of calories and harmful attributes, there are several hypothesized mechanisms for the apparent health benefits. Fruits and vegetables are high in dietary fiber which is under-consumed yet associated with a slew of benefits including protection against the development or worsening of gastrointestinal diseases.[30] The high quantities and synergistic effects of flavonoid phytochemicals[31] are also thought to contribute to high antioxidant capacity. Antioxidants defend against excess free radicals which may underlie diseases such as cancer and cardiovascular disease.[32]

Though estimates vary, the negative health impact of low fruit and vegetable consumption is profound. In 2005, up to 2.6 million deaths per year were attributed to inadequate fruit and vegetable consumption.[10] In 2013, the World Health Organization estimated that 1.7 million deaths were attributable to low fruit and vegetable consumption, as well as approximately 14% of gastrointestinal cancer, 11% of ischaemic heart disease and 9% of stroke deaths.[33] In the U.S., 20,000 cancer cases were estimated preventable by increasing fruit and vegetable consumption by one serving per day,[34] though the results were slightly less impressive in a European study that estimated a 1% reduction in cancer risk per gram increase in daily consumption.[35] Notwithstanding issues regarding the epidemiologic evidence underlying these estimates, the evidence overwhelmingly favours health benefits with little risk associated with increased fruit and vegetable consumption.

2.1. Fruit and vegetable consumption

While the health benefits are long established, most people around the world do not consume enough fruits and vegetables. Globally, only countries representing 0.4% of the world's population meet fruit and vegetable consumption targets.[9] The problem is not new. Early reports from the U.S. found 91% of the population did not meet the recommended 2-3 fruit and 3-5 vegetable servings per day in 1976-1980.[36] Fewer data on servings consumed are available in Canada,[1] but the 2004 Canadian Community Health Survey suggested that half of Canadians did not meet the recommendations for at least 5 servings of fruits and vegetables per day, and that this proportion has stayed relatively steady since the mid-2000s. Quebec consistently has the highest consumption.[37,38]

Low fruit and vegetable consumption is not for lack of awareness. By the mid-1980s most people in the U.S. were aware of the basic tenets of healthy eating messages.[39] Still, billions of dollars are spent by government agencies, non-for-profits and industry

[1] The Canadian Community Health Survey Annual Component measures frequency of consumption rather than quantity consumed

around the world to promote fruits and vegetables with inconsistent results. Rekhy et al[11] evaluated Denmark's 6-a-day campaign, Australia's 'Go for 2&5', the U.S.'s 'Fruits & Veggies – More Matters,' New Zealand's '5+ A Day' programme, and the UK's 'Food Dudes' programme. Approaches with targeted settings and age-groups were more promising than general approaches, but evaluations of general campaigns confirmed that a reported increase in awareness did not coincide with a reported increase in consumption.[13,40,41] Lack of behavior change points to a larger disconnect between awareness and action. While awareness is known to be heavily influenced by guidelines, campaigns and media, actions are less-so except in the most resourceful population subgroups.[42,43]

Increased awareness has been inadequate to change behaviours partly because this approach relies completely on the individual. Such an approach tends to assume everyone has the ability to reach some equal level of baseline healthfulness if they are aware of how to do so. The gap between awareness and practice suggests in isolation individual approaches cannot work.[15,29] Though heavily criticized,[44] the theory of planned behavior encompasses the relationship between knowledge, intentions and action. It hypothesizes that sufficient knowledge can improve intention, and thus actions. However, the original author[45] argues that the theory is less useful when people have less preferential control over their behaviours. In the extreme, behavior change may require opportunities or resources that the person simply does not have.[46]

People with lower income and education levels are the least likely to eat enough fruits and vegetables. John Boyd Orr's, *Food, Health and Income*, documented a social gradient in diets and health as early as 1936 Great Britain.[47] By the 1990s, lifestyles high in fruit and vegetable intake were already more common in non-manual social groups, those with higher education and incomes in the U.K.[48] and in the U.S.[40] Fruit and vegetable consumption remains higher in those with higher education and/or least poverty across Northern and Western European countries,[49] the U.S.[50] and Canada,[51,52] even after accounting for fulfillment of other needs according to Maslow's hierarchy of needs.[53] The gradient in fruit and vegetable consumption persists because, where they exist, trends towards healthier eating habits are confined to particular social classes.[12]

Awareness campaigns enhance these disparities when people with higher pre-existing resources and concerns about healthy eating are more responsive.[13,14]

Where the propensity for lifestyle changes is unequal or limited, interventions targeting unsupportive environments complement those targeting people directly. Environmental strategies create opportunities and remove barriers to support individuals to adopt healthier habits. Public health adopted the idea from a longer history of ecologic conceptualizations of health in social sciences where individual transactions are viewed to occur within broader social, community and policy contexts.[54] As resources and capital vary extensively between individuals, emphasizing individual choices to improve their own health at the expense of considering broader contexts is increasingly criticized.[55] Targeting environments in concert with individuals may be more effective to generate equitable change in health behaviours.[56] Interventions to improve dietary choices are not an exception.[57]

2.2. Food environments

2.2.1. Food environment dimensions

Food environments are the collective attributes of communities where we make most of our food purchasing decisions. Conceptually, they are situated in the renewed focus on environmental determinants of health, though regional barriers such as poor access and affordability of healthy foods were noted in 1986.[58] To my knowledge, the earliest paper that measured the food environment was in Hampstead, London in 1990. It demonstrated that poor districts had overall cheaper food but i) less healthy choices (ratio 1:3 healthy:unhealthy food choices while it was 1:1 in affluent areas) and ii) a bigger price differential between unhealthy and healthy baskets.[39] Current interest in measuring this concept has escalated since around the year 2000. A large body of work has characterized spatial patterning of food environments as they relate to indicators of deprivation, its association with health outcomes and, to a smaller extent, its evolution over time. Varying levels of food environment healthfulness are assumed to worsen disparities in socioeconomic disparities in consumption, and policy interventions can target food environment attributes in the hope of reducing these.

Characterizations of the food environment have evolved since their initial adaptation from a marketing model that explains consumer behavior. The original 1960 marketing model defines 4 Ps that influence purchasing decisions: product, price, promotion and placement.[59] The most widely cited conceptual model for food environments adapts these four Ps. Glanz's *community nutrition environment* includes types and locations of food outlets, or McCarthy's 'placement.' Her *consumer nutrition environment* consists of features within stores including price, promotion, placement (in-store display), and 'availability.'[15] Availability encompasses only part of McCarthy's other P, product, as it refers to the existence of (un)healthy products rather than products' features. These features may or may not meet consumer preferences. Subsequent literature has returned this P to the conceptual model in its entirety. 'Product' consists of product availability combined with some measure of product features that impact consumer preferences.[60] These preference features can include acceptability, variety and/or quality.[61,62]

My guiding framework to describe the food environment maps to the original McCarthy model, while drawing on the redefinitions offered by Glanz[15] and Charreire[61]: accessibility, affordability, availability and variety. A summary is provided in Table 2.1.

Table 2.1: Summary of popular conceptual frameworks

Original McCarthy[59]	**Original definition**	**Names within food environments literature**
Product	Products are desired by consumers	Variety*[60] Presence[60] Availability*[15] Quality[61] Acceptability[61]
Price	Products have attractive prices	Affordability*[15]
Placement	Products are located conveniently in terms of time and space	Accessibility*[15] Diversity[61] Accommodation[61]
Promotion**	Products are perceived as better (e.g. discount, display, etc.)	Promotion[15]

*Terms used in my thesis
**Not a focus of my thesis

Accessibility

Accessibility is measured as the count or density of store types in a given area. Better accessibility means the area is serviced by a higher proportion of retail food outlets that are likely to sell healthy foods like fruits and vegetables, i.e. supermarkets, supercenters, and produce stores. By definition, healthy food retailers are treated as a homogenous category. Accessibility is usually operationalized as the distance from homes to nearest stores,[62] or as an ecologic measure of interest in itself. The relative accessibility of healthy and unhealthy outlets forms the basis of the modified food retail environment index (mFREI),[63] a key indicator used by the Centre for Disease Control.[63,64] Reflecting the disproportionate research focus on accessibility, policy initiatives such as Obama's Healthy Food Financing Initiative[1] target accessibility in areas with higher deprivation indicators, which are often assumed to have poorer accessibility.

Despite this policy enthusiasm, accessibility has an unclear relationship with health outcomes. The introduction of a new supermarket has repeatedly failed to correspond with increases in fruit and vegetable consumption overall or in low income areas.[65-67] Such quasi-experiments are particularly robust, but heterogeneity between studies has otherwise limited the ability to reach overarching conclusions.[68,69] Where baseline dietary quality is low or in some socio-economically disadvantaged areas, adding supermarkets may still be helpful.[70,71] In addition, there are few studies of rural areas where new supermarkets could plausibly make a bigger difference.[68] Interestingly, accessibility is more consistently predictive of body mass index and obesity.[69,71-75] The seemingly contradictory findings mean that accessibility directly affects obesity without going through diet or reflects methodological challenges. The other food environment dimensions seem more predictive of consumption, though there is less work attempting to link them to other health outcomes.[76]

Affordability

The affordability dimension encompasses the cost of healthy foods, and the cost of healthy relative to unhealthy foods. Though affordability technically refers to cost relative to means, cost is usually studied in its own right as a proxy for affordability; the terms are used interchangeably here. Healthier diets are consistently found to be costlier.[77-79] The 2005 U.S. Dietary Guidelines market basket would require a low income family to devote 43% to 70% of their food budget to fruits and vegetables, for example.[77] The observation is not limited to the U.S. A meta-analysis of 27 studies in 10 countries suggested a healthy food-based dietary pattern is about US$1.50/day more than an overall less healthy diet.[80] Compounding the average higher cost, the healthy versus unhealthy dietary cost disparity may have grown over time,[81] though this increased disparity is not a consistent finding.[82]

Better fruit and vegetable affordability generally leads to improved consumption and health outcomes. To change dietary behaviours, a subsidy of at least 10-15% has been suggested,[83] though the maximum health benefit is thought to require taxation on unhealthy foods combined with subsidizing healthy foods.[84] Specifically, the mean percentage change in demand for a 1% change in price is estimated at 0.70 (95% CI 0.41, 0.98) and 0.58 (95% CI 0.44, 0.71) for fruits and vegetables, respectively.[85] In the limited observational literature that has linked regional affordability information and health outcomes, lower healthy food prices have resulted in less body mass index gain in lower income children,[86] and lower blood sugar in low income type 2 diabetes patients.[87] Each 10 cent increase in price per serving for fruits and vegetables also decreased the odds of having the item at home by 23% in the U.S.[88] A notable exception exploited yearly variation in prices to find no association with body weight from 1982-1996.[89]

Though cost is the most commonly reported barrier to adopting healthier diets, improving affordability is not a panacea. In 1945, George Stigler used an optimization algorithm to identify a diet of only wheat flour, dried navy beans, evaporated milk, cabbage and spinach was both affordable and nutritionally adequate.[90] However, few people would adhere. Once palatability and social acceptability are considered, the cost of a nutritionally-adequate diet rises.[91,92] For example, a French study found that the lowest cost diet was EUR 3.18/day, but its contents deviated significantly from the most popular products that French people habitually consume.[93] Even cheaper, a plan was subsequently found for < EUR 1.50/day, though this plan, and similarly low cost plans, consisted mainly of carrots, low-fat milk and organ meat (i.e liver).[94]

Availability/Variety

A healthy diet is thus not solely about nutrition, and ensuring affordability of any nutritious diet is not enough. The availability and variety dimensions conceptualize this issue. While it may be possible to meet nutritional requirements on a relatively low food budget, food has to be familiar and meet cultural, societal and personal preferences.[95] Availability refers to the presence of some basic set of popular food products. The dimension is measured either as the continuous proportion of these basic products that are available, or using a 1/0 measure based on a cut-off in the proportion. In contrast, variety refers to the range of individual items available which goes beyond whether a product is simply available or not.[96] Availability and variety are not mutually exclusive; variety can only exist among products that are available. Unless explicitly defined in the study, they can also be difficult to separate and are thus treated together here.

Like affordability, studies indicate positive effects of better availability and/or variety on consumption, but research is highly limited. In the marketing literature, greater variety is known to result in higher sales[97] up to some point where the choice overwhelms consumers.[98] Doubling the shelf-space devoted to hard fruits, for example, was associated with a 44% increase in sales in supermarkets in a seminal 1974 study.[99] In the only experimental study measuring variety in public health, hospital canteen consumers purchased more fruits and vegetables when their availability and variety increased.[100] The few observational studies are mixed. For instance, in Hartford, Connecticut, a low-income population was more likely to purchase fruits and vegetables where product variety was higher,[101] and, similarly, residents of neighbourhoods carrying at least five varieties of dark-green and orange vegetables consumed 0.17 more daily servings on average in Detroit.[102] However, others did not find an association between fruit and vegetable shelf-space and body mass index,[76] or between greater healthy food availability and consumption of fruits and vegetables.[103]

2.2.2. Food environments and disparities

Food environments and regional inequalities

Though the impact on health is of ultimate interest, understanding how food environments relate to socioeconomic disparities in fruit and vegetable consumption is an important policy focus. Deprivation amplification theory posits that the characteristics of worse-off areas amplify the effects of individual disadvantage. Individual and area factors then work together to have a deleterious effect on health.[104] Spatial patterning is sometimes of interest in its own right but food environments research is usually grounded in deprivation amplification theory. Most research has sought to determine whether food environments vary in accordance with area socio-economic characteristics. The dietary inequalities that are observed consistently over time and in different places are potentially created or worsened by poorer food environments. Policy initiatives have often assumed they are.[104]

Accessibility is not generally found to differ systematically based on area deprivation outside the U.S., putting the existence of food deserts into question. Food deserts are deprived areas with low access to healthy food retailers. In the U.S., places with high proportions of African American or low income residents are often found to be underserved by supermarkets,[105] and deprived rural areas may be particularly prone to lower accessibility.[106] But food deserts have not generally been found elsewhere, including in Canada. Food deserts have been identified in Hamilton[107] and London, Ontario,[108] but studies in Montreal,[109] Quebec City, [110] Gatineau[111], the greater Toronto area,[112] Edmonton,[113] and Saskatoon[114] have not found that deprived areas had less access to healthy food outlets. Conversely, three Canadian studies have found areas with higher deprivation are better serviced by healthy options.[111,112,114] In light of this evidence, the idea of food deserts has been described as a 'factoid,'[115] and a recent U.S. commentary suggested it was time for policy to move on.[17]

Though studied less, affordability and availability are similarly consistent across varying socioeconomic areas, but the picture seems different for variety. A form of 'poverty penalty,'[116] food is sometimes thought to be more expensive in worse areas. However this is not shown to be the case.[19,117,118] In the U.S.,[19,119] UK [92,96,120] and Australia,[118] availability is also similar between areas with different levels of deprivation. One systematic review found 19 studies were about evenly split between finding more deprived areas had lower fruit and vegetable availability or finding no difference.[117] Research is more indicative that lower variety aligns with regional deprivation. Areas with higher deprivation have less varieties of food products,[121] including low-fat, low-salt,[122] or regular versus baked chips.[19] A Southampton, UK study found that the most deprived areas had the lowest variety of healthy products,[96] though another UK study did not find a difference.[123] The gradient existed, but was less pronounced, in Australia.[118]

In addition to regional deprivation indicators, rural areas have been a subject of interest because they potentially have unique issues that could result in worse food environments. Mainly, stores are less proximal, but the effect of this is controversial. Hartley et al[124] argue that even with further stores, accessibility is less of a significant issue because people in these communities are accustomed to travelling larger distances. Further, rural residents may adopt alternate strategies including freezing, hunting and gardening, or relying on farmer's markets and orchards.[124,125] Others have confirmed some level of acceptance of travelling large distances and such adaptation strategies,[126] but not all rural residents have access to vehicles or housing that can support large freezers.[127] Further, availability[128] and variety [129] might be lower in the stores that do exist, and, where available, produce quality can be inferior.[130] Healthy items are sometimes more expensive,[129,131] but not systematically so.[120] Unique food environment dynamics and attributes in rural areas are ultimately possible,[132] though they may depend on heterogeneity between rural areas.

Food environments and time

Spatial patterning and area characteristics have dominated food environments research largely at the expense of time. The literature to date is overwhelmingly cross-sectional.[14,55,69,133] Recently, more interest in time trends has started to provide additional insight into the accessibility and affordability dimensions. Such temporal variation has the potential to worsen consumption disparities when seasonal and temporal changes disproportionately affect deprived areas and households. Dynamic food environments are a newer area of food environments research made possible by exploiting longitudinal data and more advanced methods.

Accessibility has been studied more for overall temporal trends, while season has been shown to influence both accessibility and affordability. Accessibility to healthy food retailers has improved over time, particularly in more deprived areas in the U.S.,[134] but not necessarily relative to unhealthy outlets.[135] The healthy options in more deprived areas may also be less stable.[136] Both accessibility and affordability improve during harvest seasons. Farmers' markets and roadside stands provide more options nearby for residents.[137] Some produce also becomes relatively cheaper during harvest seasons, though prices are not lower across the board.[138,139] Within supermarkets, a Montreal study found approximately a 10 cent fluctuation in the price per serving of fruits between 2008 and 2010.[139] Despite the potential for seasonal disparities in affordability, such estimates are rare. Measurement and methodologic issues have limited dynamic food environments research.

2.2.3. Measuring food environments

Geographic information systems & proxy assessments

Food environment dimensions are often indirectly measured using accessibility. Attempts to measure accessibility capitalize on the wide availability and relative ease of measurement using geographic information systems and secondary data sources such as business listings.[62] Though common, its use as a proxy for other dimensions is unsubstantiated.[62,70,140,141] Supermarkets do have cheaper healthy food prices and better variety than alternate store types such as convenience stores.[16,74,107,142] However, by definition, this has to be the case due to economies of scale; comparing within-store characteristics across store types is not particularly insightful.[117] More importantly, cross-store type comparisons do not provide evidence that more supermarkets will enhance affordability, availability or variety of healthier items. Treating all supermarkets as equal inhibits detecting important within-store differences and prevents studying them over time.[117,143] These unmeasured characteristics likely contribute to disparities in consumption.

Relying solely on accessibility also means relying on the increasingly contentious assumption that individuals shop with equal propensity at any store within given geographic boundaries.[61,144] Typically these boundaries are administrative or based on some radius around households such as 1 kilometer. The proximity assumption is not without basis. Models from urban planning suggest that people would indeed preferably shop at their nearest stores.[15] Modelling travel like Newton's theory of gravity suggests that attraction between origin and destination is proportional to distance.[145] More generally, the principal of least effort offers that the, "invariable minimum that governs all varying conduct of an individual is least effort."[146] However, this notion is losing its pertinence in food environments research. The 'local trap' falsely presumes people access all their life necessities within boundaries that researchers define as 'local.'[147] Measuring activity spaces using Global Positioning Systems[148] demonstrates that people frequently travel outside researcher-defined 'local' boundaries. Instead, they seek out food stores that meet their preferences.[149] [150,151]

In-store assessments

Instead of using accessibility as a proxy, affordability, availability and variety are most often measured directly using in-store food audits. The most common audit tool is the Nutrition Environment Measures Survey-Stores (NEMS-S).[152] It consists of a list of 10 popular fruit and vegetable products for which prices can be checked and a binary or proportion-available measure can be used to derive availability. It was rooted in Glanz's model which did not distinguish between availability and variety.[15] More recently, responding to the need for a more comprehensive measure than the NEMS-S, Black et al developed an additional in-store tool that specifically included a variety construct defined as, "*The number of different choices within a product range based on: product flavour, product size, fair trade/organic range or no-name/low-cost range.*"[123] More often authors have created their own tools on a per-study basis. Such checklists or tools have had between 7 and 80 food products.[117] Rarely, affordability has been measured using local food price indices,[62] though these are not store-specific. The lack of studies directly measuring within-store characteristics at the food outlet level is a large gap in the literature.[69]

Directly measuring within-store dimensions will enable more insight into disparities in the static and dynamic food environments that people experience. Within-store characteristics drive choices about where and when to shop and what to buy. Low income households will tend to shop at places with lower food prices,[123,144] and are more sensitive to price fluctuations.[153] As such, assessing accessibility to low price chains is more indicative of where low income households are likely to shop. Two novel studies make use of this fact. In the UK, dietary quality was lower in individuals who lived the furthest away from their income-based economically-matched supermarket.[154] In the U.S. those shopping at low price supermarkets had higher body mass index.[151,155] Links to health outcomes, however, are limited by the dearth of literature characterizing within-store dimensions. So far, stratifying supermarkets according to price levels has revealed lower accessibility to low price stores[140] and that those who shop at low price stores face a lower fruit and vegetable variety.[123] The extent to which other drivers, such as chains, might matter to the association between accessibility and affordability is unclear. Further, the differences in availability and variety in different chains and temporal and seasonal variation in affordability are rarely investigated.

Though they carry the potential to help understand the influence of food environments on consumption disparities, a central challenge in assessing within-store dimensions remains the in-store measures. The cost and time required to visit and survey stores naturally limits spatio-temporal resolution.[156] In the early stages of food environment research, Glanz (2005) developed the Nutrition Environment Measures Survey to address the need for a valid measure of the food environment.[15] Some have undertaken a massive effort to run such in-store surveys across large geographic spaces, for example, by visiting 466 stores in the UK.[92] While these studies can provide an unparalleled level of detail about in-store environments, repeating this effort elsewhere on an annual or seasonal basis is not sustainable. These methods pose a barrier to expanding the evidence base as it relates to large spatial scales, chains with different price levels and time.

2.3. Summary

Fruits and vegetables are under-consumed despite their substantial health benefits. While campaigns have successfully increased general awareness and occasionally led to behavior changes, they disproportionately exclude those with lower levels of income and education. These groups may face environmental hindrances that preclude incorporating more fruits and vegetables in their diet despite knowing that they should. Over the last two decades, we have seen substantial progress in learning about the potential impact of food environments on consumption disparities through an explosion in the number of articles, systematic reviews, reviews of reviews, and policy documents.

The food environment can be characterized in terms of its accessibility, affordability, availability and variety. The literature and corresponding policy initiatives have focused heavily on supermarket accessibility. Due to the relative ease of data collection, the accessibility dimension has also been used frequently as a proxy for other food environment dimensions. However, the relation between accessibility and area deprivation and dietary outcomes is inconsistent and the use of accessibility as a proxy is unsubstantiated. Without direct measurement, store-based, spatial and temporal heterogeneity in affordability, availability and variety are masked because all supermarkets are treated as equal. As these in-store dimensions influence purchasing decisions, the suppressed heterogeneity may worsen disparities in consumption. The reliance on in-store surveys to measure affordability, availability and variety has inherently limited the development of policy-relevant evidence.

Chapter 3 Thesis Aims

The manuscript-based thesis contains three studies that each addresses a specific objective. Addressing gaps in the food environments literature to date as highlighted in Chapter 2, I examine the understudied food environment dimensions of affordability, availability and variety over space and time.

For all of my objectives, I use six years (2008-2013) of weekly digital grocery store transaction data obtained from a large marketing company to overcome the problems and limitations associated with using in-store tools. These data are at the individual product level, with information on store, chain, date, price, sales and region, for a total of 10,435,888 records pertaining to fruit and vegetable sales in Quebec, Canada. The sampling frame included stores with at least $2 million in revenue, and all major chains were represented. The strata to select the random sample of stores consisted of census metropolitan areas, smaller cities and towns and rural areas and towns with population less than 5000.

Using the transaction data, I aim to improve upon work over the last two decades by developing evidence that will contribute to a more resilient and equitable food environment. This evidence will support designing policies beyond those that target food deserts, to help address the structural determinants of poor dietary habits and, ultimately, obesity, cancer and chronic disease.

In Chapter 4, I determine the relationship between regional affordability and accessibility. This objective challenges the commonly held assumption that having more supermarkets, or other food retailers considered healthy, means greater affordability. This assumption underlies much of the policy interventions formulated to date. While some quasi-experimental evidence has emerged in the last two years, the relationship between accessibility and affordability has not been described on a large spatial scale.

In Chapter 5, I estimate the disparities in availability and variety across chain affordability levels. I also determine whether these disparities depend on regional income. In general, very few studies have explored availability and variety despite their theorization in food environment conceptual models, so this study contributes to developing research in this area. Further, I test the hypotheses that those shopping in lower price stores and/or in poorer areas may have access to a lower variety of fruits and vegetables. This could limit purchasing and consumption.

In Chapter 6, I measure the seasonal and temporal variation in regional fruit and vegetable affordability. Few studies have been able to consider the time dimension of food environments. Seasonality has not been previously studied despite potentially posing barriers to affordability. In addition to quantifying seasonal price differences for fruits and vegetables, I determine whether these differences occur to a greater extent in higher price regions. Seasonal price fluctuations could compound pre-existing lack of affordability.

Across the studies in this thesis, I also aimed to prioritize data management and reproducibility. These were a challenge in this thesis because of the multitude of large data sources and levels of processing required. In the Supplementary Information, I describe the approach to organizing the data and provide a link to the code used to analyse the data in all three papers.

Chapter 4

Study One: The dynamic between affordability and accessibility of healthy food

My first objective is to estimate the expected change in the regional average price of fruits and vegetables with an additional supermarket. I overcome past limitations to directly measuring affordability by using digital grocery store transaction data, which are available for all of Quebec. Merging these data with business directory data by region enables affordability and accessibility to be determined for each region, and the estimation of their association. I controlled for regional socio-demographic characteristics using information from the 2006 Canadian census and 2011 National Household Survey, and also used this information to measure the same associations stratified by rural and urban regions to determine whether it was different.

This study contributes to questioning the heavy policy focus of incentivizing more healthy food retailers to improve the food environment. In the literature, accessibility indicators are often used as a proxy for affordability because of the difficulty to obtain price data. As such, policy intervention has often assumed that improved accessibility, or more supermarkets, will make healthy food more affordable to those who might otherwise face difficulty in purchasing it. While supermarkets do tend to offer cheaper prices than grocery stores or convenience stores, this premise does not imply more supermarkets means better regional affordability. A more nuanced understanding of the drivers of the association between accessibility and affordability could better inform future interventions to make healthy foods more easily accessed by all socioeconomic groups.

4.1 Abstract

Introduction

Affordability is an important barrier to consuming more fruits and vegetables. Regions with more supermarkets are often assumed to have healthy foods at lower average prices than regions with fewer supermarkets. Previous work has not considered the dynamic between accessibility and affordability in relation to other regional characteristics.

Methods

We used weekly data on digital transactions and business directory data to measure affordability and accessibility in 47 regions in Quebec, Canada (2009-2013). To measure affordability, we determined the mean weekly price of a standard basket of fruits and vegetables in each region. To measure accessibility, we determined the proportion of supermarkets relative to unhealthy outlets, and relative to population size. We estimated the association between accessibility and affordability using a linear model of the log of price with random intercepts for region, and fixed effects for chain, year, and month. We additionally stratified the models by rural and urban areas.

Results

The average supermarket density was 0.11 per 1000 people, and the average proportion of supermarkets to unhealthy outlets was 0.09. The mean serving price for fruits and vegetables was 39.3 (SD = 2.46) and 40.6 (SD = 4.8) 2010 Canadian cents, respectively. We found the supermarket density was positively associated with cost for vegetables (percent price difference = 8.0 [95% 2.0; 15]), but other accessibility indicators were not associated with prices. In rural areas, the association between price and accessibility was consistently negative, but in urban areas the opposite trend was observed. Chains were the most important predictor of price.

Conclusion

Regional characteristics and chain-make up are important determinants of the dynamic between accessibility and regional fruit and vegetable affordability. Public health

interventions based on incentivizing supermarket construction alone should not be expected to improve regional affordability of fruits and vegetables.

4.2 Main Text

4.2.1 Background

Average fruit and vegetable consumption remains below recommended levels in most high-income countries.[1] Affordability is the most commonly reported barrier to consuming healthier foods including fruits and vegetables.[2-4] A specific threshold for affordability is subject to interpretation but according to one U.S. study, a low income family would need to devote 43% to 70% of their food budget to achieving a diet with the recommended amount of fruits and vegetables.[5] While particularly problematic in low income populations, cost is a concern even among those who are relatively food secure.[6] The healthy versus unhealthy dietary cost disparity may have also grown over time,[7,8] potentially making healthy food less economically viable for more people.

A large body of research has aimed to understand differences in affordability across neighbourhoods, stores and time, often as a feature of the food environment. This environment creates local opportunities and barriers to consuming healthy foods such as fruits and vegetables. Glanz[9] describes the food environment as consisting of the *community nutrition environment* and the *consumer nutrition environment*. The former includes types and locations of food outlets, also called accessibility. The *consumer nutrition environment* consists of features within stores such as price, herein affordability.[9] Such ecological models of health behaviour have increased in popularity because approaches that have relied entirely on the individual have failed to reduce the prevalence of risk factors for obesity and chronic disease. Food environments research helps to inform potential structural approaches to improving population diet.

Measuring affordability usually relies on in-store surveys of product prices while accessibility can be measured without visiting stores. Measuring accessibility uses methods relying on geocoded business listings to determine the residential proximity or density of (un)healthy food outlets in geographic areas. One approach is to measure healthy food options relative to unhealthy food options, however dichotomized, while the other is to measure healthy food retailers per capita.[10] The former is commonly operationalized as the modified food retail environment index (mFREI), or the number of outlets such as supermarkets and produce stores divided by the total number of outlets including fast food and convenience stores.[11,12] From a resource-perspective, measuring accessibility is relatively simple. However, measuring affordability is more resource intensive because it typically requires internal store audits using a checklist or market-basket tool.[13,14]

Presuming supermarkets are more affordable than other types of stores has helped to overcome the difficulty in obtaining price information. But evidence for this assumption is equivocal. Theoretically, accessibility is used as a proxy for affordability[15] because supermarkets have a cheaper and greater variety of healthy food options relative to convenience or small grocery stores.[16] However, such cross-store type comparisons do not imply that adding any supermarket to any area will render the region's healthy food options more affordable. Supermarkets are not a homogenous category. Chains and stores make decisions to locate and market differently depending on zoning, the socio-demographic makeup of a region, and local competition.[17] These characteristics may also differ in rural regions,[18,19] where adding a supermarket may have a different impact. Given the substantial policy interest in incentivizing the addition of healthy food retailers,[20] supermarket nuances require further clarification to determine the extent to which all supermarkets are created equal.

In this study, we determine the association between regional accessibility and average regional affordability of fruits and vegetables using digital grocery store scanner data from supermarkets in the province of Quebec, Canada. In doing so, we challenge the common assumption that supermarkets inherently improve affordability.

4.2.2 **Methods**

Study Design

Automatically collected digital grocery store scanner data provide an opportunity to understand food prices in relation to accessibility across a large geographic and time scale. Though traditionally used for marketing purposes, scanner data can be analyzed to characterize places where people are less able to purchase healthy foods due to factors such as cost [21]. We acquired these data from a large marketing firm, and they form the basis for the affordability measure in this ecologic study. sampling frame included all chain supermarkets, pharmacies, convenience stores, and gas stations that had more than \$1 million or \$2 million in sales for convenience stores and supermarkets, respectively. These data excluded all frozen and canned fruit and vegetable products, and do not provide sufficient detail on store-brand products for analyses. Stores were selected by the marketing firm using a stratified random sample based on the degree of rurality of cities and towns (metropolitan cities, smaller towns and rural areas).

All measures were derived at the regional health service authority level (i.e. the *réseaux locaux de services* or RLS). We restricted the analysis to regions containing sampled stores, so that data on food price would be available (n=57/96). Of these regions, we also excluded those with more than 10% of weeks with missing data. This study included the period January 2009 to December 2013 (inclusive).

Fruit and vegetable affordability

Our affordability measure was the weighted mean weekly price of a standard basket of fruits and vegetables. The standard basket had 29 and 23 commonly consumed vegetable and fruit products, respectively (Section 4.6, Table 4.11). We standardized individual items to price per 100 gram or 140 gram serving for vegetables and fruits, respectively, deflated to average 2010 Canadian dollars. Monthly market weights were generated by determining the market share by volume sold of each product. The mean weighted price thus placed greatest weight on the most popular products. The least and most expensive items within each product category were truncated, as is a common practice.[22] These truncation points were determined by removing the top and bottom 5th percentile; the top and bottom 1st percentile was truncated instead in a sensitivity analysis.

Fruit and vegetable accessibility

The full details are available in Section 4.3. Briefly, we calculated the mFREI and the supermarket density per 1000 population to measure accessibility for each region. Food outlets were classified using the Expanded Points of Interest Files (DMTI Spatial Inc Enhanced Point of Interest Markham: DMTI Spatial Inc, 2008-2013) The data were first narrowed to seven standard industry classification codes (SIC) representing outlets that would likely be selling food, and we undertook further efforts to improve the accuracy of the supermarket category due to the potential for misclassification.[23] Through a validation sub-study, we found that approximately 84% of stores classified as supermarkets were in fact convenience (n=278/1330), small grocery/corner stores (n=659/1330), or unidentifiable and thus removed. The small grocery/corner stores were often indistinguishable from convenience stores and were thus excluded from the mFREI in the main analysis. The outlets in the remaining years were re-classified according the positive predictive values identified through the sub-study.

Regional characteristics

Regional socio-economic and population data were obtained from the 2006 Canadian Census and 2011 National Household Surveys, administered by Statistics Canada.[24] Socio-economic status was measured as the Material Deprivation Index which uses information on education, income and rurality to classify areas into deprivation quintiles.[25] We also extracted population density per square metre and the Census-defined rural zone, as described below. Values between 2006 and 2010, and 2011 and 2013 were imputed using a simple linear regression model to obtain continuous measures for years 2009-2013 because this was the time frame for our affordability data.

Statistical Analysis

We used linear random effects regression to model the association of each accessibility indicator with the affordability of fruits and vegetables separately, for a total of four models. To account for the various levels of correlated data, we included fixed effects for the chain and random intercepts for each region. The fixed effects were intended to account for non-time varying characteristics of each chain that could affect fruit and vegetable prices, such as the fact that some chains have cheaper or more expensive prices on average. The random effects reflect our assumption that the regions are a random sample of a larger population of regions in Quebec, and allow a different baseline effect for each region. We also adjusted for regional material deprivation index, rural zone, population density per square metre, and year, and allowed for a flexible seasonal pattern by including a 4-knot spline for month. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used to assess model fit, and Moran's I statistic[2] was used to assess residual spatial auto-correlation in the error terms. The significance of the statistic was based on a permutation-based p-value with 1000 simulations, using p=0.05 as a cut-off.

We then stratified by degree of rurality. As recommended by Statistics Canada, we used defined the most rural zone as having no influence (primarily measured by commuting zones) from population centres of 10,000 or more people, while weakly-influenced and moderately-influenced zones are considered less rural. Population centres are considered as Zone 1. To categorize the data, we considered regions with average zone values one to three as non-rural and regions with average zone values greater than three as rural.

In Section 4.4, we provide the results of several additional analyses to support the robustness of our findings. As large supermarkets are more accurately identified through business directories than smaller grocery stores,[23,26] we re-estimated the primary model after restricting supermarkets to chains only. We also ran the same models including the smaller grocery stores that were excluded from the main analysis. Additionally, we determined the impact of removing outlier regions identified by assessing region-specific intercepts and the impact of truncating food products at the 1st and 99th percentile instead of the 5th and 95th percentile. Finally, we repeated our analysis using the forward sortation area, a smaller but less socially meaningful geographic boundary, to assess the sensitivity of our findings to the decision to aggregate at the RLS level.

All main analyses were done in *R* (R Development Core Team), while data were managed in PostgreSQL.

4.2.3 **Results**

Of the 57 regions with price information, 47 had 10% or fewer weeks with missing data and were thus available for analysis. Overall, the number of food outlets was consistent over time (Table 4.1). The average number of supermarkets per region was 25 (SD =40.4). There were 203 stores with available price information in our sample. There were 17 rural regions and 30 urban regions that differed on a variety of characteristics. Rural regions were on average more deprived, had fewer supermarkets and had lower populations, though the mFREI and supermarket density per 1000 people were similar (Table 4.2).

Table 4.1: Total numbers of food outlets, accessibility and affordability indicators

Year	Supercentres, Supermarkets	Convenience, Fast Food, Pharmacy	mFREI	Supermarket Density/ 1000 people	Fruits[i]	Vegetables[i]
2009	863	8230	0.1	0.11	38.64	41.05
2010	827	7961	0.09	0.1	39.05	40.69
2011	840	8069	0.1	0.1	37.97	40.19
2012	885	8563	0.09	0.11	38.65	41.56
2013	886	8563	0.09	0.11	39.79	42.39
i. Refers to the mean serving price in cents ($Cdn 2010)						

Table 4.2: Average characteristics in urban and rural regions

	Urban	Rural
Regions	30	17
Population	193549	68758
Deprivation Quintile	2.8	3.7
Supermarkets	19.3	7.2
Supermarket Density	0.1	0.1
mFREI	0.097	0.093

The best model to estimate the association between accessibility and affordability included month, year and chain, with random intercepts for region. The largest improvement in model fit occurred after including the dummy variables for chain. The incorporation of this information had a considerable impact on the AIC and BIC, suggesting chain substantially contributed to correctly modelling price information. There was minimal improvement in model fit after including the other regional characteristics (Section 4.6 Table 4.12).

Using this model, the associations between accessibility and affordability were consistently negative, except for the association between supermarket density and the mean price of the vegetable standard basket (Table 4.3). Comparing two similar regions, one with an additional supermarket per 1000 people is expected to have an average price 8 (95% CI 2, 15) percent higher for vegetable products. In other words, we expect the serving price to be an average of 41.2 cents versus 38 cents in the region with the additional supermarket. The remaining associations were not substantively important. For example, a ten percent increase in the proportion of healthy stores relative to unhealthy stores is expected to be associated with an average fruit serving price decrease of 0.9% (95% CI 2, 0).

Table 4.3: Regression results for accessibility versus affordability

	Fruits[i] **[mean serving price range = 32 to 44 cents]**	**Vegetables[i]** **[mean serving price range = 32 to 49 cents]**
Supermarket Density[ii] [95% CI]	-5 [-11;0]	8 [2;15]
mFREI[iii] [95% CI]	-0.9 [-1.6;-0.3]	-1.37 [-2.13;-0.6]
i. Linear regression model adjusted for year, year2, a 4-knot spline for month, chain and random intercepts for region ii The coefficients refer to the percent change in mean serving price associated with one additional supermarket per 1000 people iii The coefficients refer to the percent change in mean serving price associated with a 10% increase in the number of supermarkets relative to unhealthy food stores		

We did not find evidence of spatial autocorrelation between the residuals, suggesting the errors were independent after adjusting for the covariates. The Moran's I^2 statistic p-value ranged from 0.39 to 0.59 for the four models (Section 4.6 Table 4.13).

The association between accessibility and affordability was different in the urban and rural stratum (Table 4.4). In rural regions, better accessibility was associated with improved affordability, but the opposite was true in urban areas. The coefficients were stronger in magnitude for supermarket density than for mFREI, as in the primary analysis.

Table 4.4: Results stratified by urban and rural areas

	Urban[i]	**Rural[i]**
Supermarket Density[ii]		
Vegetables	26.3 [18.0;34.8]	-2.5 [-11.2;6.3]
Fruits	5 [-1;12]	-19 [-27;-11]
mFREI[iii]		
Vegetables	0.4 [-0.6;1.5]	-2.3 [-3.3;-1.2]
Fruits	0 [-1;0]	-2 [-2.9;-1.1]
i. Linear regression model adjusted for year, $year^2$, a 4-knot spline for month,		

chain and random intercepts for region
ii The coefficients refer to the percent change in mean serving price associated with one additional supermarket per 1000 people
iii The coefficients refer to the percent change in mean serving price associated with a 10% increase in the number of supermarkets relative to unhealthy food stores

Further exploration revealed that the high positive coefficient for the effect of supermarket density in urban areas was primarily driven by a positive association between supermarket density and affordability in three regions. On average, these regions had 12.08 (95% CI 11.63, 12.53) fewer supermarkets than the rest of the urban regions. They were also more expensive (mean difference per vegetable serving in cents = 2.62 [95% CI 2.05, 3.18]).

Restricting the analysis to the supermarkets identifiable as chain supermarkets, including small grocery stores, removing outlier and alternate price truncation did not impact the results. In addition, an analysis using forward sortation areas, a smaller geographic unit, instead of RLS did not alter the results. Full results are available in Section 4.4 Additional Analysis.

4.2.4 **Discussion**

We found that improved accessibility is not consistently associated with more affordable fruits and vegetables. In contrast to previous work that has relied on in-store surveys or proxies, we directly measured affordability by using automatically collected digital grocery store transaction data. Using these data enabled estimation of prices at a sample of individual stores across a large geographic space. While the presence of supermarkets is often assumed to improve regional affordability of healthy foods, we found the dynamic depends on the regional characteristics and the specific chains.

Our disparate findings in rural and urban areas confirm the importance of pre-existing regional characteristics that could determine the impact of adding more supermarkets. Rural areas may have a unique dynamic due to the presence or lack of certain chain stores, larger travel distances and item choice limitations.[27-29] The associations between accessibility and affordability were more in the expected direction, i.e. more supermarkets per capita or relative to unhealthy outlets were associated with lower average regional prices. Conversely, accessibility was either not associated or associated with lower affordability (higher costs) in urban areas. Where at least a few fruit and vegetable sources exist, it is possible that additional sources may not have a substantial impact on the local food environment,[30] or prices increase due to greater variety, for example baby versus regular carrots. Increased variety can introduce less affordable items,[31] i.e. baby carrots are more expensive per serving than regular carrots. In addition, the entry of a new supermarket may encourage existing stores to differentiate further, for example, by introducing more expensive items or additional services that increase costs to avoid direct competition.[32] In our sample, urban regions with generally higher prices and fewer supermarkets had particularly positive associations between supermarket density and cost, suggesting greater adjustment.

Chain make-up is an important driver in the dynamic between accessibility and affordability. In addition to a significant improvement in model fit, the precision of our estimates increased and the magnitude changed once we included chain variables in the model. This suggests that *which* supermarkets are present is more important to driving regional prices than the mere presence of supermarkets. Others have found that price [33] and the variety of products offered[34] are different depending on the supermarket. Presuming that "all supermarkets offer the same variety of healthy foods at the same cost to consumers,"[27] is thus not consistent with these or previous findings.

Our results corroborate literature that has challenged the presumption that increasing numbers of supermarkets imply increased affordability.[27] Supermarkets are cheaper than convenience or small grocery stores,[16] but introducing supermarkets where at least some already exist does not necessarily have positive effects on the local food environment. Most recently, researchers in the U.S. found that opening a new supermarket did not impact fruit and vegetable availability or prices in a food desert, though the average junk food price increased.[30] Cummins *et al*[35] arrived at a similar conclusion in Glasgow, Scotland.[35] Both researchers commented that whether people use the new supermarket is another issue, as habit and familiarity may drive purchasing more than convenience. A new supermarket may not have any impact on the accessibility or affordability people experience as a result.

Limitations

We had to exclude store-brand products, frozen and canned fruit and vegetable products. Store-brand products represented an average 9% market share in our data, however, the estimates would only be biased if, for example, a high-end chain sold frozen produce at prices cheaper than a low-end (budget) chain. We also did not have price information for non-supermarkets. Though supermarkets are the main source of food shopping for most of the population, supermarket average prices may not reflect the average price of fruits and vegetables consumed outside the home or acquired through other sources such as growing or community gardens. Our choice of RLS as the primary area of interest may have resulted in misclassification of exact store locations because the forward sortation areas at which store price data was available do not always map exactly to within the RLS. We mitigated this by also estimating effects at the forward sortation area unit level in an additional analysis; however, aggregating the data to other geographic boundaries may affect these results.

4.2.5 **Conclusion**

Public health researchers and practitioners have tended to assume that more physical opportunities to access healthy foods will ultimately improve consumption. This assumption would be supported if regions with greater accessibility tend to have more affordable healthy food, due to the importance of affordability for consumption. Using digital grocery store transaction data, we studied the dynamic between the accessibility of supermarkets and the affordability of fruits and vegetables. While most studies have assumed this relationship is positive, we found that at a regional scale, greater accessibility was not consistently associated with better affordability. The relationship between accessibility and affordability depended critically on the presence of certain chains, and the direction and magnitude of the association differed when considering fruits versus vegetables and rural versus urban areas. As a result, interventions to improve access by incentivizing the development of supermarkets may have counterintuitive effects unless they consider the specific characteristics of the region.

4.3 Defining Accessibility

Introduction

Food outlets used to defined accessibility in the main analysis presented in Section 4.2 were classified using the Expanded Points of Interest Files (DMTI Spatial Inc Enhanced Point of Interest Markham: DMTI Spatial Inc, 2008-2013) The data were first narrowed to seven standard industry classification codes (SIC) representing outlets that would likely be selling food (Table 4.5).

Table 4.5: Standard Industry Classification (SIC) codes used to narrow the business directory

SIC Code	Description	Example
5411	Grocery Stores	Provigo©
5311	Department Stores	Zellers©
5141	Groceries, General Line	Costco Wholesale©
5812	Eating places	McDonald's©
5912	Drug Stores and Proprietary Stores	Uniprix©
5431	Fruit and Vegetable Markets	Le jardin mobile ©
5541	Gasoline Service Stations	Petro-Canada©

A key word index was created to search store names within the SIC categories and classify them according to store types and chain name. Five store types were used: fast food, pharmacy, supermarket, superstore, produce and convenience store.

The key store name index corresponding to these categories was determined based on general knowledge of the Quebec food environment, grey sources such as lists of chains obtained from Wikipedia and pharmacy database, and additional Google searches for lists of store names. On top of specific chains, I also included terms that would indicate that the outlet belonged to a particular category, for example 'pizza,' or 'casse-croute,' though these were limited compared to the chain key words. The key terms lists ultimately consisted of 81 key terms for fast food, 112 key terms for supermarkets, 27 for convenience stores (including gas stations), 11 for pharmacies, and 3 key terms for supercenters. The produce category was based on the '5431 (Fruit and Vegetable Markets)' SIC code only, as otherwise it was not possible to generate a key terms list that was comprehensive enough to capture small-scale produce stores such as 'fruiteries'. This SIC code was excluded from the analysis.

Due to the well-known inaccuracy of business directory data, in addition to the above, one year of data (2008) were manually cleaned to identify incorrectly classified stores that could affect our results. In particular, I was concerned about incorrectly classified stores that should/should not belong to the numerator of accessibility measures, i.e. healthier stores. As such I focused on correcting the 'supermarket' category above. For this category, there were three principle issues that created the necessity for the manual cleaning: i) ambiguous store names resulting in misclassification of chain supermarkets; ii) supermarkets that were actually convenience stores; iii) lack of distinction between smaller grocery stores and larger supermarkets.

Methods for business directory cleaning

There were 1959 supermarkets in the 2008 business directory after removing duplicates. Of these, 1344 could not be matched to a known supermarket chain and were thus ambiguous. These primarily consisted of stores with 'marché', 'épicerie' or 'alimentation' in their names (1330/1344). A 10% random sample of each of these was drawn. If no supermarkets were found, the entire category was reclassified as either small grocery or convenience stores depending on the true classifications of stores in the random sample. If true supermarkets were identified, the entire category was verified and reclassified as appropriate.

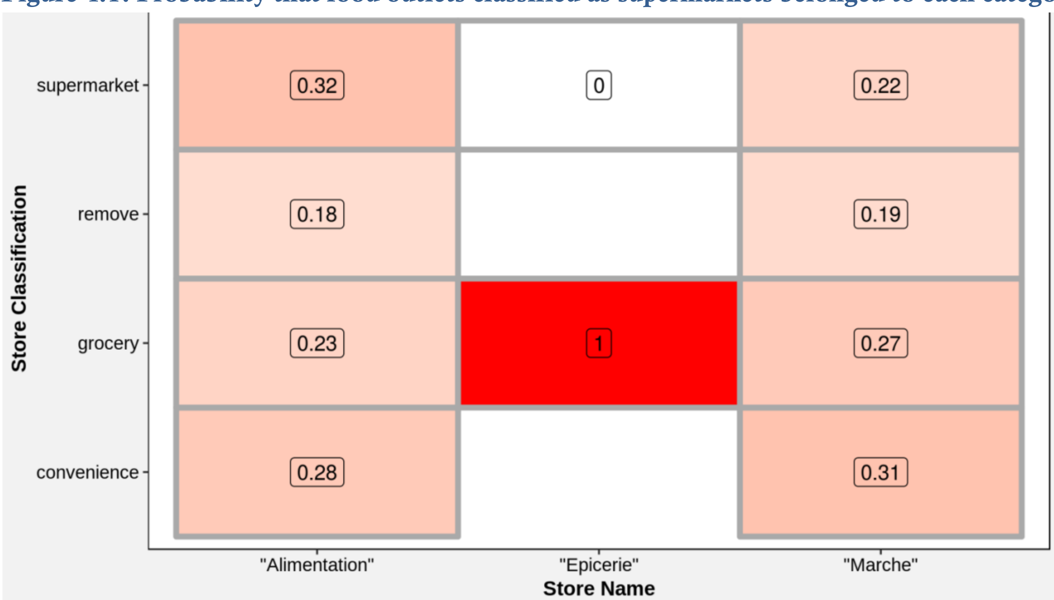
I used external resources to verify whether each store was a supermarket. First, I visited the stores virtually using Google eye view with a time lapse to the correct year. As the visits were not always sufficient due to limited available years, inaccurate address or inability to find the store (particularly in rural areas), websites, Facebook pages, phone numbers, customer reviews and photos were used as additional evidence depending on what was available. Based on the verification the outlet retained its supermarket classification with the addition of a chain name if missing, or it was classified as a small grocery store or convenience store. The small grocery versus convenience store classification was ambiguous and only based on the store size and perceived product content where available. Therefore, small grocery stores were excluded from the primary analysis. If there was insufficient evidence to suggest that the store existed, it was removed. Service stations were considered as convenience stores.

Once the data were re-classified, the probability that a store classified as a supermarket was actually a supermarket was derived overall and for each name category. While there was substantial overlap in the stores over the years that enabled propagating the reclassification with certainty, the positive predictive values were applied to supermarkets that were present in 2009-2013 but not in 2008. Supermarkets were thus reclassified as convenience stores or grocery stores in accordance with the requisite probabilities derived from the 2008 data.

Results of business directory cleaning

Of the 1330 that were ambiguous, 1109 outlets were classified as either grocery (n=659), convenience (n=278), or removed (n=172). This left 221 (16.62%) as supermarkets (Figure 4.1). No true supermarkets were identified in the 'épicerie' category, so all 418 outlets were classified as grocery stores based on only a 10% random sample. The remaining two categories were fully verified.

Figure 4.1: Probability that food outlets classified as supermarkets belonged to each category



Applying these probabilities resulted in a decrease in the number of stores classified as supermarkets and an increase in stores classified as convenience and small grocery stores. The grocery store category was classified as 'healthy' in the numerator for the derivation of the modified retail food environment index in sensitivity analysis. I expect that the new numbers better reflect the true number of supermarkets in the province. (Table 4.6)

Table 4.6: Totals before and after classification corrections

no_convenience	no_supermarket	no_grocery	source
21548	3951	2811	Corrected
21019	6517	2022	Original

4.4 Additional Analyses

4.4.1 Restrict supermarket category to chain only

When I restricted to chain supermarkets only, the associations remained consistent. Those for mFREI and affordability grew stronger in magnitude (Table 4.7)

Table 4.7: Restrict to chain supermarkets

	Main Analysis	Restrict to Chains
Fruit-density	-0.05 [-0.11;0]	-0.07 [-0.13;0]
Fruit-mFREI	-0.009 [-0.016;-0.003]	-0.01 [-0.017;-0.002]
Veg-density	0.08 [0.02;0.15]	0.06 [-0.03;0.14]
Veg-mFREI	-0.0137 [-0.0213;-0.006]	-0.0215 [-0.0308;-0.0122]

4.4.2 Include the grocery category

Including the grocery category as part of the numerator for the mFREI did not impact the direction of our results, and the estimates were similar in magnitude (Table 4.8).

Table 4.8: Include small grocery store category in mFREI numerator

	Main Analysis	Include Grocery Stores
Fruit-mFREI	-0.009 [-0.016;-0.003]	-0.007 [-0.013;-0.002]
Veg-mFREI	-0.0137 [-0.0213;-0.006]	-0.0235 [-0.0309;-0.0162]

4.4.3 Remove outlier regions

Outliers were identified by plotting the region-specific deviations from the overall model intercept. From these, I identified two outlier regions that deviated more than expected relative to other regions. The main analysis was repeated without these regions, and was consistent (Table 4.9).

Table 4.9: Exclude outliers

	Main Analysis	Exclude Outliers
Fruit-density	-0.05 [-0.11;0]	-0.04 [-0.09;0.01]
Fruit-mFREI	-0.009 [-0.016;-0.003]	-0.007 [-0.013;-0.001]
Veg-density	0.08 [0.02;0.15]	0.08 [0.02;0.14]
Veg-mFREI	-0.0137 [-0.0213;-0.006]	-0.0127 [-0.0202;-0.0052]

4.4.4 Alternate price truncation

In the primary analysis, I truncated the top and bottom most expensive items within each product category at the top and bottom 5th percentile. In this robustness check, the top and bottom 1% was used (Table 4.10). The results were consistent.

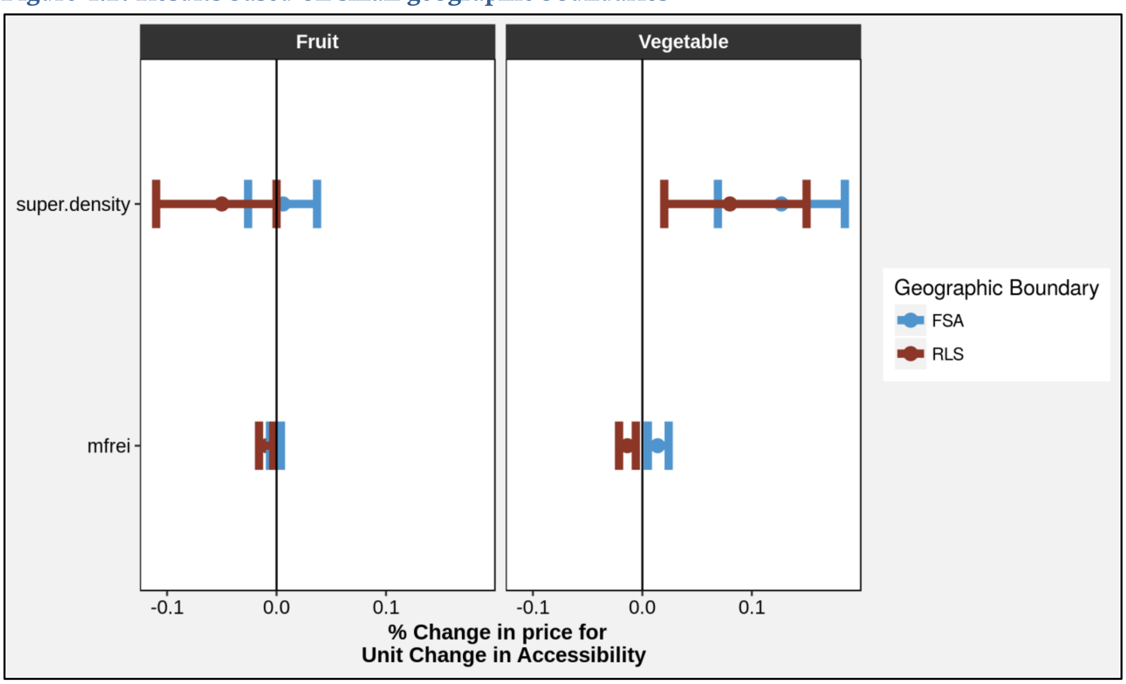
Table 4.10: Alternative price truncation

	Main Analysis	Alternate Truncation
Fruit-density	-0.05 [-0.11;0]	-0.04 [-0.09;0.01]
Fruit-mFREI	-0.009 [-0.016;-0.003]	-0.01 [-0.016;-0.004]
Veg-density	0.08 [0.02;0.15]	0.05 [-0.01;0.12]
Veg-mFREI	-0.0137 [-0.0213;-0.006]	-0.0083 [-0.0161;-5e-04]

4.4.5 Alternate geographic boundary

In the primary analysis, I used the RLS geographic boundaries to define regions. This choice of boundary may have influenced our results. Here, I define region as 'forward sortation area' (FSA), which are units used by the postal service that are usually smaller geographic areas then the RLS. In total there were 121 FSAs. The mean number of sampled supermarkets per FSA is 1.93 because the majority of FSAs had only one or two sampled stores, in contrast to the RLS where almost all regions had at least two and usually more. To account for the non-independence of the measurements with in an FSA, we used robust standard errors were clustered at the FSA level. Overall, the results were similar to our main findings (Figure 4.2). Though the mfREI coefficient for vegetables was positive, it remained substantively close to zero.

Figure 4.2: Results based on small geographic boundaries



4.5 References

1. Micha R, Khatibzadeh S, Shi P, Andrews KG, Engell RE, Mozaffarian D. Global, regional and national consumption of major food groups in 1990 and 2010: a systematic analysis including 266 country-specific nutrition surveys worldwide. *BMJ Open.* 2015;5(9).

2. John JH, Ziebland S. Reported barriers to eating more fruit and vegetables before and after participation in a randomized controlled trial: a qualitative study. *Health Education Research.* 2004;19(2):165-174.

3. Haynes-Maslow L. A qualitative study of perceived barriers to fruit and vegetable consumption among low-income populations, North Carolina, 2011. *Preventing Chronic Disease.* 2013;10.

4. Dijkstra SC, Neter JE, van Stralen MM, et al. The role of perceived barriers in explaining socio-economic status differences in adherence to the fruit, vegetable and fish guidelines in older adults: a mediation study. *Public Health Nutrition.* 2015;18(05):797-808.

5. Cassady D, Jetter KM, Culp J. Is price a barrier to eating more fruits and vegetables for low-income families? *Journal of the American Dietetic Association.* 107(11):1909-1915.

6. Mook K, Laraia BA, Oddo VM, Jones-Smith JC. Food security status and barriers to fruit and vegetable consumption in two economically deprived communities of Oakland, California, 2013-2014. *Preventing Chronic Disease.* 2016;13:E21.

7. Monsivais P, McLain J, Drewnowski A. The rising disparity in the price of healthful foods: 2004–2008. *Food Policy.* 2010;35(6):514-520.

8. Powell LM, Chriqui JF, Khan T, Wada R, Chaloupka FJ. Assessing the potential effectiveness of food and beverage taxes and subsidies for improving public health: a systematic review of prices, demand and body weight outcomes. *Obesity Reviews.* 2013;14(2):110-128.

9. Glanz K, Sallis JF, Saelens BE, Frank LD. Healthy nutrition environments: concepts and measures. *Am J Health Promot.* May-Jun 2005;19(5):330-333.

10. Beaulac J, Kristjansson E, Cummins S. A systematic review of food deserts, 1966-2007. *Preventing Chronic Disease.* 2009;6(3):A105.

11. Greer S, Schieb L, Schwartz G, Onufrak S, Park S. Association of the neighborhood retail food environment with sodium and potassium intake among US adults. *Preventing Chronic Disease.* 2014;11:E70.

12. Centres for Disease Control and Prevention. Children's Food Environment State Indicator Report. In: Health and Human Services, ed 2011.

13. 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):S124-S133.

14. Thornton LE, Kavanagh AM. Understanding the local food environment and obesity. In: Witten K, ed. *Geographies of obesity environmental understandings of the obesity epidemic*. New York: Routledge; 2010.

15. Ver Ploeg M, Dutko P, Breneman V. Measuring food access and food deserts for policy purposes. *Applied Economic Perspectives and Policy.* 2014;37(2):205-225.

16. Cummins S, Smith DM, Aitken Z, et al. Neighbourhood deprivation and the price and availability of fruit and vegetables in Scotland. *Journal of human nutrition and.* Oct 2010;23(5):494-501.

17. Black JL, Carpiano RM, Fleming S, Lauster N. Exploring the distribution of food stores in British Columbia: associations with neighbourhood socio-demographic factors and urban form. *Health & Place.* 2011;17(4):961-970.

18. Byker Shanks C, Ahmed S, Smith T, et al. Availability, price, and quality of fruits and vegetables in 12 rural Montana counties, 2014. *Preventing Chronic Disease.* 2015;12:E128.

19. Hardin-Fanning F, Rayens MK. Food cost disparities in rural communities. *Health Promot Pract.* May 2015;16(3):383-391.

20. Flournoy R. A healthy food financing initiative: An innovative approach to improve health and spark economic development. Reinvestment Fund, PolicyLink, CA, US. 2010.

21. Ma Y, Labban A, Cherian M, Shaban-Nejad A, Buckeridge DL, Dubé L. System of indicators for the nutritional quality of marketing and food environment: product quality, availability, affordability, and promotion. *Diet Quality*; 2013:383-396.

22. Čížek P. General trimmed estimation: robust approach to nonlinear and limited dependent variable models. *Econometric Theory*. 2008;24(6):1500-1529.

23. Liese AD, Barnes TL, Lamichhane AP, Hibbert JD, Colabianchi N, Lawson AB. Characterizing the food retail environment: impact of count, type, and geospatial error in 2 secondary data sources. *J Nutr Educ Behav*. Sep-Oct 2013;45(5):435-442.

24. Statistics Canada. Taking a Census of Population. 2006. http://www12.statcan.ca/census-recensement/2006/ref/dict/overview-apercu/pop2-eng.cfm#06. Accessed 22-02-2016.

25. Pampalon R, Raymond G. A deprivation index for health and welfare planning in Quebec. *Chronic Diseases and Injuries in Canada*. 2000;21(3):104.

26. Paquet C, Daniel M, Kestens Y, Léger K, Gauvin L. Field validation of listings of food stores and commercial physical activity establishments from secondary data. *The International Journal of Behavioral Nutrition and Physical Activity*. 2008;5:58-58.

27. Jiao J, Moudon AV, Ulmer J, Hurvitz PM, Drewnowski A. How to Identify Food Deserts: Measuring Physical and Economic Access to Supermarkets in King County, Washington. *American Journal of Public Health*. 2012;102(10):e32-e39.

28. Hendrickson D, Smith C, Eikenberry N. Fruit and vegetable access in four low-income food deserts communities in Minnesota. *Agriculture and Human Values*. 2006;23(3):371-383.

29. Walker RE, Keane CR, Burke JG. Disparities and access to healthy food in the United States: A review of food deserts literature. *Health & Place*. 2010;16(5):876-884.

30. Ghosh-Dastidar M, Hunter G, Collins RL, et al. Does opening a supermarket in a food desert change the food environment? *Health & Place*. 2017;46:249-256.

31. Dunn RA, Sharkey JR, Lotade-Manje J, Bouhlal Y, Nayga RM. Socio-economic status, racial composition and the affordability of fresh fruits and vegetables in neighborhoods of a large rural region in Texas. *Nutrition Journal*. 2011;10(1):6.

32. Kusum L. Ailawadi, Jie Zhang, Aradhna Krishna, Kruger MW. When Wal-Mart enters: How incumbent retailers react and how this affects their sales outcomes. *Journal of Marketing Research*. 2010;47(4):577-593.

33. Mackenbach JD, Burgoine T, Lakerveld J, et al. Accessibility and affordability of supermarkets: associations with the DASH diet. *Am J Prev Med*. 2017;53(1):55-62.

34. Black C, Ntani G, Inskip H, et al. Measuring the healthfulness of food retail stores: variations by store type and neighbourhood deprivation. *International Journal of Behavioral Nutrition and Physical Activity*. May 23 2014;11(1):69.

35. Cummins S, Findlay A, Petticrew M, Sparks L. Retail-led regeneration and store-switching behaviour. *Journal of Retailing and Consumer Services*. 2008;15(4):288-295.

4.6 Supplemental Tables

Table 4.11: Fruit and vegetable standard baskets

Type	Products
Fruits	apple; apricot; banana; grapes; cherry; orange; peach; pear; nectarine; melon; watermelon; avocado; papaya; plum; pineapple; blackberry; blueberry; strawberry; raspberry; kiwi; grapefruit; mango; pomegranate
Vegetables	tomato; turnip; vegmix; brusselsprouts; carrot; cauliflower; asparagus; celery; cucumber; potato; lettuce; onion; squash; sprouts; radish; beet; broccoli; cabbage; peas; pepper; endive; mushroom; beans; leek; bokchoy; parsnip; sweetpotato; zucchini; eggplant

Table 4.12: All model regression results

Supermarket Density – Vegetable*						
Model**	**Coefficient****	**low.ci**	**high.ci**	**aic**	**bic**	**Model description**
1	0.006	-0.112	0.123	-19577.9	-19527.8	Year only
2	0.007	-0.106	0.121	-21757.5	-21674	M1 + month
3	0.083	0.02	0.146	-58569.2	-58418.9	M2 + chain
4	0.086	0.023	0.149	-58550.8	-58375.5	M3 + material deprivation + rural zone + population
mFREI - Vegetable						
model	**coefficient**	**low.ci**	**high.ci**	**aic**	**bic**	
1	-0.023	-0.038	-0.009	-19582.8	-19532.7	Year only
2	-0.023	-0.037	-0.009	-21762.7	-21679.2	M1 + month
3	-0.014	-0.021	-0.006	-58564.8	-58414.6	M2 + chain
4	-0.013	-0.021	-0.006	-58544	-58368.7	M3 + material deprivation + rural zone + population
Supermarket Density – Fruits						
model	**coefficient**	**low.ci**	**high.ci**	**aic**	**bic**	
1	-0.1	-0.196	-0.003	-31625.7	-31575.7	Year only
2	-0.081	-0.167	0.004	-39272.1	-39188.6	M1 + month
3	-0.053	-0.105	-0.001	-67600.5	-67450.2	M2 + chain
4	-0.056	-0.109	-0.004	-67562.1	-67386.8	M3 + material deprivation + rural zone + population
mFREI - Fruits						
model	**coefficient**	**low.ci**	**high.ci**	**aic**	**bic**	
1	-0.015	-0.027	-0.004	-31623.5	-31573.4	Year only
2	-0.013	-0.024	-0.003	-39270.2	-39186.7	M1 + month
3	-0.009	-0.016	-0.003	-67600.7	-67450.4	M2 + chain
4	-0.01	-0.016	-0.003	-67562.4	-67387.1	M3 + material deprivation + rural zone + population

*Outcome = log(mean serving price)
**All models include random intercepts for region

***Coefficient = coefficient on accessibility indicator, representing the percentage difference in mean serving price

Table 4.13: Assessment of spatial autocorrelation in primary model residuals

	Moran I statistic	p.value
Fruit-mFREI	-0.052	0.59
Fruit-Supermarket Density	-0.05	0.592
Veg-mFREI	0.009	0.388
Veg-Supermarket Density	-0.017	0.459

Chapter 5 Study Two: Disparities in the availability and variety of fruits and vegetables

In this objective, I estimate the disparity in the availability and variety of fruits and vegetables between chain price levels (high, medium and low) and between stores situated in low versus high income areas. The availability and variety measures are constructed at the level of store-month over five years to allow the assessment of changes within stores over time. Similar to previous research, availability is operationalized as the presence of a fruit and vegetable standard basket, and variety as the number of items within each fruit and vegetable category. In contrast to most previous research, I use price levels within chains as a proxy for where low income consumers are more likely to shop instead of assuming that consumers shop primarily in their local area of residence.

The results from this study will advance the limited literature that measures the availability and variety dimensions of the food environment. Further, they will identify differences between supermarkets that people with different budgets tend to frequent. While past work has studied these dimensions using in-store surveys, the use of digital transaction data in the current work allows assessment over larger geographical and temporal scales. A better understanding of the availability and variety of fruits and vegetables in supermarkets can reveal the extent to which consumer preferences may or may not be met, which can in turn impact purchasing. As in other parts of this thesis, the current study challenges the assumption made frequently in the literature and in policy rationale that supermarkets are a homogenous category.

5.1 Abstract

Objective

To quantify disparities in the availability and variety of fruits and vegetables.

Methods

We use monthly digital grocery store transaction data (2009-2013) for 185 supermarkets in 9 chains in Quebec, Canada. The primary exposure was chain affordability: low, medium or high price level. Availability was defined as the presence of a standard basket of fruits and vegetable products. Variety was defined as the number of varieties of each product. The disparity in availability was assessed using a linear model and disparity in variety was assessed using a quasi-Poisson model. Explanatory variables included chain affordability, area income level, and interaction terms. We used cluster-robust standard errors.

Results

Relative to high price chains, the standard basket was 5.6% (95% CI 2.6;8.5) less likely to be available in medium-price chains, and 7.2% (95% CI 3.7;10.6) less likely to be available in low-price chains. Medium and low-price chains also had 13% (95% CI 8.3; 18.7) and 21% (95% CI 18.1; 24.1) fewer varieties per fruit and vegetable product. Availability and variety was similar in regions with different income levels.

Conclusion

The lower availability and variety of fruits and vegetables at more affordable supermarkets creates disparities in the selection of these foods, which can reduce opportunities for healthy eating.

5.2 Main Text

5.2.1 Background

Public health interest in the food environment has grown over the last two decades. The most common metric of the food environment, and target of policy interventions,[1] is accessibility, or the number of healthy food retailers in an area. Recently, the focus on accessibility has been criticized because this metric assumes all healthy food retailers are the same, and because the metric is not consistently associated with improved diet or health outcomes.[2] Thus, other dimensions are increasingly being explored. While affordability is known to be important, products must also be palatable, familiar and meet personal preferences.[3] For example, when treated as an algorithmic optimization problem, nutritionally-adequate, low cost diets consist mainly of carrots, low-fat milk and organ meat.[4] Evidently, few people would adhere to this diet. Availability and variety are the food environment dimensions that correspond to product selection.

Availability is the presence of popular food products while variety encompasses the product features that match consumer preferences. Initially defined in Glanz's 2005 conceptual model of the food environment,[5] availability has most often been measured using the Nutrition Environment Measures Survey-Store (NEMS-S).[6] This tool contains a list of 10 popular fruit and vegetable products for which a binary or proportion-available measure can be used to derive availability. The NEMS-S does not assess fruit and vegetable variety however, so another in-store tool that includes variety was developed in 2014. This tool includes a count of varieties based on the number of different choices within a product range. The choices reflect different flavours, organic versus regular, brands, formats, and sizes.[7] Availability and variety are not mutually exclusive; variety can only exist among products that are available.

Better availability and variety are thought to improve consumption, though public health evidence is limited. The association between availability, variety and sales has emerged from marketing. Product selection is instrumental to driving store and product choice,[8-10] and increased shelf space devoted to items increases overall sales of the product category.[11,12] In public health, studies are emerging. Two of four studies identified in a 2012 systematic review found higher fruit and vegetable availability and/or variety measured using the NEMS-S or similar checklists[13] improved consumption. Subsequently, higher product variety in stores in Hartford, Connecticut was associated with an increased likelihood of purchasing that fruit/vegetable product in a low income community.[14] This public health literature is fairly new,[13] and while it shows promise, the association between availability, variety and consumption (usually proxied by sales) is more established in marketing.

Consumers who shop at lower-priced stores or at stores in lower-priced areas may face systematic disadvantage in the availability and variety of fruits and vegetables. Previously, cheaper stores have been found to have lower availability and/or variety than higher price supermarkets.[7,15,16] Further, stores located in poorer areas may have lower availability or variety in the U.S. [15] and the U.K.[16] but the findings are inconsistent. [7,17,18] In the U.S., supermarkets in poor areas missed 0.4 more items of a healthy food basket on average relative to wealthier areas, while in the U.K. there was a 34% lower chance of having high fruit and vegetable variety in the most versus least deprived areas.[16] In Australia, the social gradient was present but less pronounced.[17,18] The evidence leans towards true differentials in variety, and to a lesser extent availability, by store type and neighbourhood socioeconomic status, but studies are still limited.

In this study, we use five years of monthly digital transactions data on fruit and vegetable purchases from supermarkets in Quebec, Canada to assess disparities in availability and variety. We determine disparities in availability and variety across chain affordability levels and between stores situated in regions with different income levels. In doing so, we overcome resource limitations associated with using in-store tools, and contribute to developing evidence on food environment dimensions which have remained relatively unexplored.[7,13,19]

5.2.2 **Methods**

Data sources

The primary data source was digital scanner data from a major marketing company. The data are transaction records that are collected automatically at cash registers in retail food stores. We use monthly data from January 2009 to December 2013 (inclusive) from a cluster-stratified random sample of stores across Quebec, Canada. Data from

2008 were used to calibrate market weights and these data were subsequently excluded from analysis. The sampling frame includes all food stores reporting at least $2 million in sales. All major supermarket chains are represented. The fruit and vegetable products measured excluded all frozen and canned fruit and vegetables products. Further, we excluded store-brand products because those data lacked sufficient detail to protect chain confidentiality. We additionally excluded dried fruits and vegetables and herbs.

The regions used were forward sortation areas (FSA), which are created for use by the postal service. Information on population density and median family income were obtained from the 2006 Canadian Census and 2011 National Household Survey.[20] A simple linear model was used to extrapolate regional population and median family income to each year between the two measured points.

Standard Basket

To develop a standard basket, we first created 'bins', or high-level groupings of fruits and vegetables. The groupings were designed to contain products that consumers would consider similar. Similar products were identified based on their names and, when necessary, their stock-keeping units. For example, different types of apples were just 'apple,' and different types of oranges were just 'orange'. Bins were ranked according to total sales dollars and total volume of sales (servings). The top half of bins in each ranking was retained. We compared the findings to the list of products identified in the NEMS-S.

Outcome measures

An availability of 1 was assigned to store-months where 85% of the standard basket was present. We varied this threshold in an additional analysis to determine the impact on our findings (Section 5.3). Variety was defined as the number of items per standard basket bin. This variation encompassed primarily different formats, brands, styles, levels of packaging and processing, reflecting previous ways of defining variety.[16,21] We excluded stores with less than 12 months of data from the availability and variety analyses, and bins with only one variety from the variety analysis.

Exposure measures

Supermarket chains were classified as ‘high’, ‘medium’ and ‘low’ price to distinguish expensive, in-between and discount style supermarkets. The classification was based on the weighted average real price-per-serving of the standard basket between 2009 and 2013 in Canadian dollars, standardized to the 2010 consumer price index for fruits and vegetables. The price-to-weight ratio of each package or unit was converted to cents per 100 gram or 140 gram serving for vegetables and fruits, respectively. The weights were the bin’s monthly servings sold relative to all sales in 2008, such that the price of popular products contributed more to the overall mean price. The weights were derived monthly to reflect seasonality. Box plots were used to visually assess the disparities and categorize chains according to their standard basket price. We then used a linear model adjusted for year to estimate the average price difference between the chain price categories.

Analysis

To model availability, we used a linear regression model with a 1/0 indicator for the availability of the standard basket. To model variety, we used a quasi-Poisson regression model with the number of items per bin as the outcome. For both, the main explanatory variables of interest were the chain price level (three-levels) and regional median household income (<=$60,000 vs > $60,000). All chains were present in regions with both income levels. We introduced an interaction term between regional income and chain price level to determine whether the association of chain price with availability/variety depended on regional income. The reference categories were the high-price category and the high-income bracket. The final models included a flexible 4-knot term to adjust for time (year-month index) and a term for the log of population density.

We accounted for heteroscedasticity and clustering due to repeated measures in the same stores by using robust standard errors clustered at the store-year level. Further, we estimated and derived standard errors for the marginal effect of chain price level on the number of bins and items using a 2-level bootstrap technique that preserves the structure of hierarchical data.[22] Analyses were mainly done in *R* (R Development Core

Team), though we used PostgreSQL to manage the data and conduct some of the analysis.

5.2.3 **Results**

Overall, there were 9 chains in the sample, representing all major supermarket chains in Quebec. Each chain had an average of 20.56 (SD=7.58) stores. After removing herbs, dried fruits and vegetables, and non-fruit and vegetable products, there were 2424 vegetable items and 2619 fruit items that were grouped into 84 and 67 bins, respectively (Section 5.5 Figure 5.3). Seventeen of the 185 stores had less than 12 months of data and were excluded.

The standard basket included 34/84 vegetable and 26/67 fruit bins. Thirteen of the sixty bins were excluded from the variety analysis because they had less than two varieties (eggplant, cherry, turnip, cantaloupe, corn, pomegranate, apricot, rappini, brussel sprouts, sweet potato, rutabaga, parsnip, plantain). All vegetable and fruit products in the NEMS-S were also identified by our approach. (Section 5.5 Figure 5.4) There was an average of 314.3 (SD=85.39) unique standard basket items per store in each month, and 5.0% (SD=4.0) were organic. Within a given store, month and chain, bins had an average of 5.4 (SD=5.75) items each. Bins with the most varieties were apple, potato, tomato, onion, and orange, which had 26.32, 17.6, 16.6, 15.8, and 15.09 varieties on average, respectively.

The high, medium and low price chain groups contained 2, 3 and 4 chains respectively. The corresponding average serving prices for the standard basket were 45.3 (SD = 1.10), 41.7 (SD = 1.83) and 33.6 (SD=1.6) cents. The proportion of standard basket bins available, the total number of items, and the average number of varieties per bin decreased with chain price levels (Table 5.1). The high, medium and low-level groups offered an average of 5.87, 5.34 and 4.92 unique items per bin, respectively. The gradient was in the same direction but less pronounced between median regional income levels (Section 5.5 Table 5.4).

Table 5.1: Availability and variety by chain price

Chain price	Available[i] (SD)	Total items (SD)	Items per bin[ii] (SD)	Stores
High	0.95 (0.22)	329.11 (63.68)	5.87 (5.94)	56

Medium	0.92 (0.27)	301.02 (94.06)	5.34 (6.01)	62
Low	0.89 (0.32)	271.38 (46.91)	4.92 (5.08)	50

i. Average proportion of the standard basket available stores in the respective income stratum
ii. Number of varieties of each fruit and vegetable product (bin)

There was a lower probability of the standard basket being available in lower and medium price chains relative to high price chains. Low price chains had an approximately 7.5 % (95% CI 4.7, 10.2) lower probability of having the full standard basket, which converted to 4.47 (2.68, 5.88)) fewer bins. The difference in availability between medium and high price chains was smaller (probability difference = -0.04 [95% CI -0.07, -0.02], or, 2.53 [95% CI 1.16, 3.87] bins). We did not find that availability varied across stores depending on regional income level (Table 5.2).

Table 5.2: Results for availability analysis

Model	Medium Price Chain[1] (Ref: High price chains)	Low Price Chain[1] (Ref: High price chains)	Low income[1] (Ref: High income)	Medium price *Low income[1] (Ref: High price/income)	Low price*Low income[1] (Ref: High price/income)
1*	-4.21 (-6.5;-1.9)	-7.45 (-10.2;-4.7)			
2**			-1.34 (-3.3;0.6)		
3***	-5.55 (-8.5;-2.6)	-7.17 (-10.6;-3.7)	-2.11 (-4.8;0.6)	2.5 (-1.8;6.8)	-0.5 (-5.8;4.8)

1. Coefficients represent mean percent difference in probability that the standard basket is availability (95% CI)
* Linear model with chain price point adjusted for 4-knot month-year index only, with robust standard errors
** Linear model with Income level adjusted for 4-knot month-year index only, with robust standard errors
*** Linear model with Income level, chain price point and interaction term adjusted for 4-knot month-year index, with robust standard errors

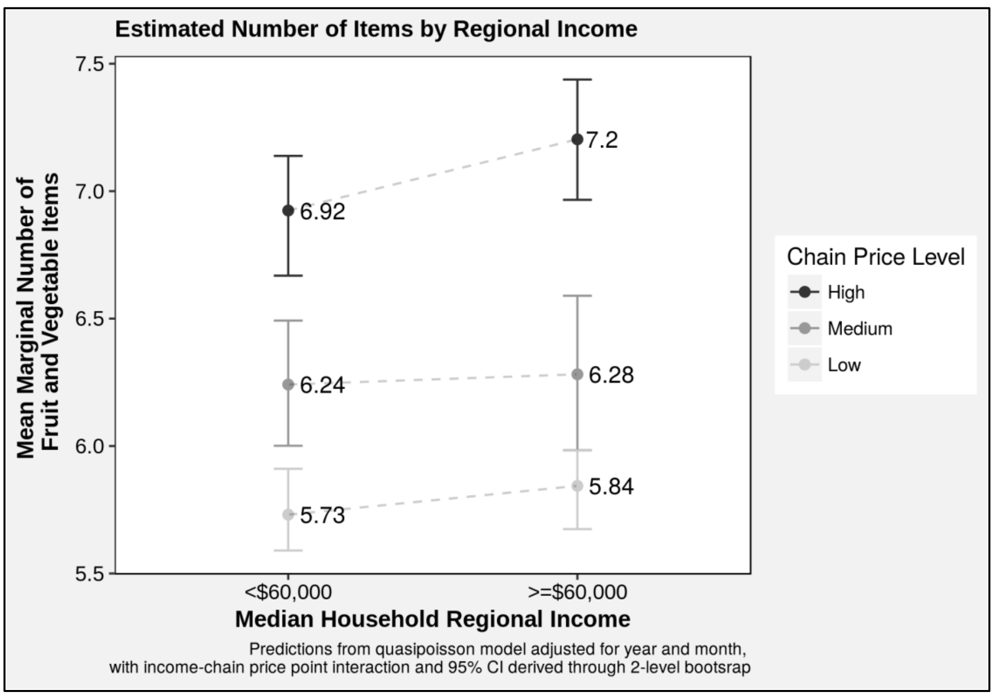
Variety followed a gradient similar to that for availability (Table 5.3). Compared to high price chains, there were 9 % (95% CI 6, 12) fewer items per bin in medium price chains, and 18% (95% CI 15, 20) fewer items in low price chains, on average. These percent differences corresponded to an average of 7.04 (95% CI 6.84, 7.25), 6.27 (95% CI 6.05, 6.51), and 5.78 (95% CI 5.67, 5.92) items per bin in high, medium and low prices chains, respectively (Section 5.5 Table 5.5). Variety was lower in stores situated in low income areas, though the effects were small in magnitude. The interaction terms in the model revealed that the difference in variety between high and low income areas was similar within chain price levels (Figure 5.1). The full model results are available in Section 5.5 Tables 5.6 and 5.7.

Table 5.3: Results for variety analysis

Model	Medium Price Chain[1] (Ref: High price chains)	Low Price Chain[1] (Ref: High price chains)	Low income[1] (Ref: High income)	Medium price *Low income[1] (Ref: High price/High income)	Low price *Low income[1] (Ref: High price/High income)
1*	-9.02 (-12.4;-5.6)	-17.69 (-20;-15.4)			
2**			-2.25 (-5.1;0.6)		
3***	-13.46 (-18.7;-8.3)	-21.11 (-24.1;-18.1)	-3.99 (-7.2;-0.8)	3.5 (-3;10)	2.24 (-1.9;6.4)

1. Coefficients represent mean percent differences in the mean number of items per bin (95% CI)
* Quasipoisson model with chain price point adjusted for 4-knot month-year index only, with robust standard errors
** Quasipoisson model with Income level adjusted for 4-knot month-year index only, with robust standard errors
*** Quasipoisson model with Income level, chain price point and interaction term adjusted for 4-knot month-year index, with robust standard errors

Figure 5.1: Variety (items per bin) by income level and chain price



5.2.4 Discussion

We analysed the availability and variety of fruits and vegetables in chains with different price points and by regional median household income. Lower price chains had lower availability and variety as compared to higher price chains. However, we found that availability and variety were similar across areas with lower and higher income. The heterogeneity we observed in availability and variety between supermarkets chains challenges the common assumption that all supermarkets offer the same selection of healthy products at similar prices.[23] Chain affordability may determine availability and variety, and exacerbate dietary inequalities as a result.

Chain prices are valuable to assess socio-economic disparities in the absence of individual-level information on shopping practices. Affordability, habit and product selection are the primary drivers of supermarket choice.[9] Product assortment, a synonym for availability/variety as used in this study, is sometimes even more influential than price. In research on food environments, people are assumed to be mainly influenced by food outlets near their residence.[24] For example, residents in areas with fewer supermarkets are assumed to have less access to healthy food. However, consumers can travel considerable distance to shop at stores and chains that meet their budget or other preferences.[25,26] Specifically, low income consumers will generally shop at lower price chains. Therefore, where the interest is in socioeconomic disparities, examining food environment dimensions across levels of chain affordability may more closely approximate that target construct than examining food environments around geography alone.

The lower availability and variety observed in low price chains can enhance disparities in the healthfulness of local food environments. Our findings corroborate others that demonstrate lower price supermarkets have less availability and variety of healthy foods.[15,16] One study found that 100% of mainstream supermarkets had at least three varieties of healthy products, while only 50% of discount supermarkets did.[16] The magnitude of our results is smaller, which is expected because our chain price categorization was more nuanced and our sample size enabled adjusting for additional characteristics. However, the similarity in the direction suggests that the phenomenon occurs outside the U.S. and the U.K. which have dominated food environments research. Up to a point,[27] better availability and variety can lead to improved consumption. As a result, the systematic disadvantage in availability and variety can perpetuate dietary inequalities based on product selection at consumers' supermarket of choice.

Our findings support the growing literature that questions regional income as dictating systematically less unhealthy food environments. Grounded in the deprivation amplification theory, the disadvantage associated with living in lower income regions is assumed to be compounded by, for example, poorer food environments.[2] The quality of produce is more consistently demonstrated to be lower in lower income areas.[28,29] However evidence for other socioeconomic food environment disparities is inconsistent. For example, the existence and relative importance of food deserts to contributing to poor diets is questionable, despite the popularity of the concept.[21,30] In Canada, several studies have found that low income areas have better supermarket access,[31-33] which may be because wealthier communities rally against commercial development.[34] Like others,[16,17,35,36] our study indicated that variety may be lower in stores in low income areas, but the effects were substantively small. Low income areas are not synonymous with poorer food environments in Quebec.

Public health interventions related to food environments have largely focused on encouraging supermarkets to locate in underserved areas. These policies improve spatial access while ignoring in-store characteristics.[37] As the first paper to explore fruit and vegetable availability and variety over a large geographic and time scale, we find that more affordable chains may create poorer food environments. Interventions that incentivize healthy food retailers to locate in low income areas may not improve dietary inequalities when the within-store contents fail to meet consumer preferences. In addition to recognizing the importance of product selection, further work to assess the necessary level of variety within fruit and vegetable categories to improve consumption would help to inform possible interventions.

Limitations

The study uses sales data, so an alternate interpretation of our findings is that the observed differences reflect differences in people's purchasing habits instead of supply-side variation. However, the item would only need to be sold once in one month to create a sales record. We used item names to measure variety, which is an imperfect proxy for item variety as items could be quite similar but still have different names. Further, we focused on the product assortment angle of variety in supermarkets (e.g. red versus green apple).[37] Thus we could not assess variety in relation to cultural acceptability (e.g. the presence of culturally-specific items beyond the set of popular items),[38] quality beyond organic versus not (e.g. bruised items),[19] or account for the fact that lower variety may mean higher quality if a store focuses on only fresh, seasonal goods. Finally, we used FSA as our regional boundary which can be heterogeneous in themselves. Household-level data linked to supermarket locations could improve estimation of income-related differences in food environment characteristics.

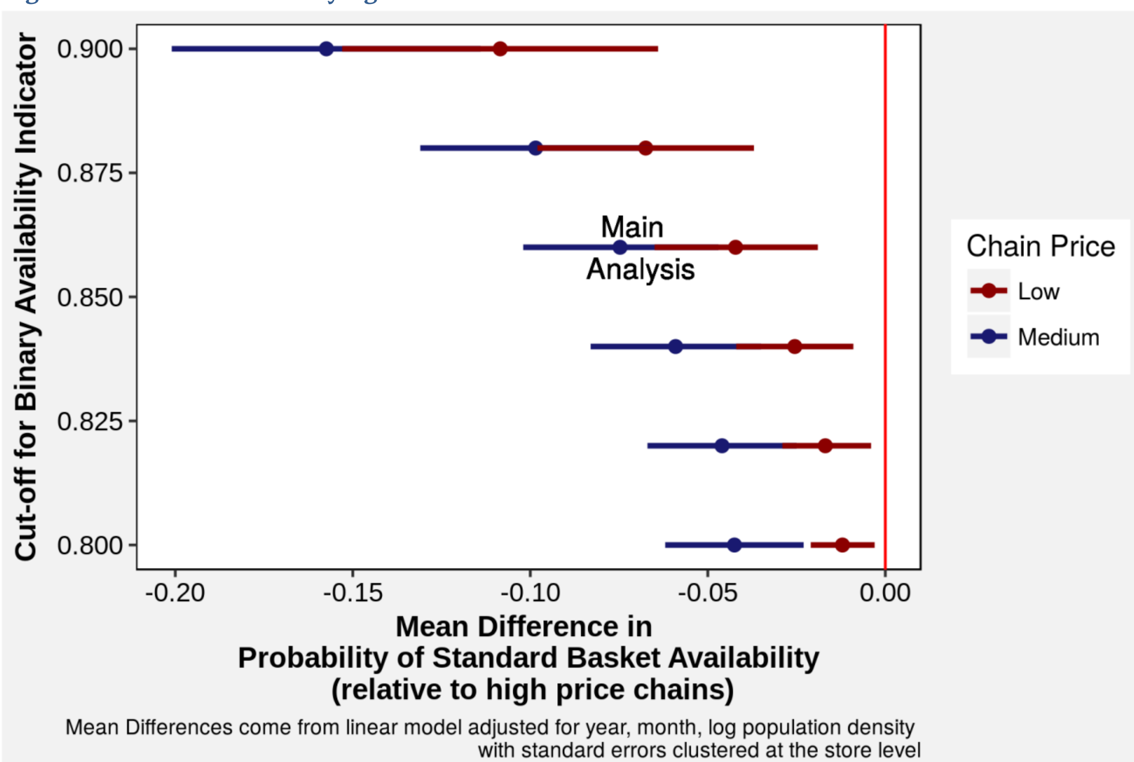
5.2.5 Conclusion

Amidst growing public health interest in the food environment, the availability and variety of healthy foods have been theorized as important dimensions. But there are few studies assessing these factors. We used digital grocery store transaction data to assess availability and variety, based on the average cost of a standard basket of fruits and vegetables at different chain supermarkets and regional income levels. Lower price chains had lower availability of the standard basket, and fewer varieties of fruits and vegetables. This study contributes to the growing literature suggesting that policy strategies that treat all supermarkets as equal may be overlooking the importance of within-supermarket differences that drive consumption.

5.3 Additional Analyses

In the main text (Section 5.2), store-months where 85% of the standard basket was present took a value of 1 for availability. In this analysis, the 85% cut-off was varied to determine the impact on the results. Like in the main analysis, I used a linear regression model with a 1/0 indicator for the availability of the standard basket as the outcome. The standard errors were clustered at the store level. The results were consistent across different cut-offs (Figure 5.2).

Figure 5.2: Results with varying income cut-offs



5.4 References

1. U.S. Treasury Department Office of Public Affairs. Obama Administration Details Healthy Food Financing Initiative. 2010. https://www.cdc.gov/chronicdisease/recovery/PDF/Healthy_Food_Financing_release.pdf. Accessed April 1, 2018
2. Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? International Journal of Behavioral Nutrition and Physical Activity. 2007;4(1):32.
3. Ragaert P, Verbeke W, Devlieghere F, Debevere J. Consumer perception and choice of minimally processed vegetables and packaged fruit. Food Quality and Preferences. 2004;15(3):259-70.
4. Maillot M, Darmon N, Drewnowski A. Are the lowest-cost healthful food plans culturally and socially acceptable? Public Health Nutrition. Aug 2010;13(8):1178-1185.
5. Glanz K, Sallis JF, Saelens BE, Frank LD. Healthy nutrition environments: concepts and measures. Am J Health Promot. May-Jun 2005;19(5):330-333.
6. Glanz K, Sallis JF, Saelens BE, Frank LD. Nutrition Environment Measures Survey in Stores (NEMS-S). Am J Prev Med. 2007;32(4):282-289.
7. Black C, Ntani G, Inskip H, et al. Measuring the healthfulness of food retail stores: variations by store type and neighbourhood deprivation. International Journal of Behavioral Nutrition and Physical Activity. 2014;11(1):69.
8. FMCG and Retail. Understanding the top drivers behind shoppers' store choices. 10-11-2016; http://www.nielsen.com/us/en/insights/news/2016/understanding-the-top-drivers-behind-shoppers-store-choices.html. Accessed June 1,2018
9. Carpenter JM, M,. Consumer demographics, store attributes, and retail format choice in the US grocery market. International Journal of Retail & Distribution Management. 2006;34(6):434-452.
10. Pollard J, Kirk SFL, Cade JE. Factors affecting food choice in relation to fruit and vegetable intake: a review. Nutrition Research Reviews. 2013;15(2):373-387.
11. Curhan RC. The effects of merchandising and temporary promotional activities on the sales of fresh fruits and vegetables in supermarkets. Journal of Marketing Research. 1974;11(3):286-294.
12. Cox KK. The effect of shelf space upon sales of branded products. Journal of Marketing Research. 1970;7(1):55-58.
13. Gustafson A, Hankins S, Jilcott S. Measures of the consumer food store environment: A systematic review of the evidence 2000–2011. Journal of Community Health. 2012;37(4):897-911.
14. Martin KS, Havens E, Boyle KE, et al. If you stock it, will they buy it? Healthy food availability and customer purchasing behaviour within corner stores in Hartford, CT, USA. Public Health Nutrition. Oct 2012;15(10):1973-1978.
15. Andrews M, Kantor LS, Lino M, Ripplinger D. Using USDA's Thrifty Food Plan to assess food availability and affordability. Food Review. 2001;24(2).
16. Black C, Ntani G, Kenny R, et al. Variety and quality of healthy foods differ according to neighbourhood deprivation. Health & Place. Nov 2012;18(6):1292-1299.
17. Ball K, Timperio A, Crawford D. Neighbourhood socioeconomic inequalities in food access and affordability. Health & Place. 2009;15(2):578-585.
18. Winkler E, Turrell G, Patterson C. Does living in a disadvantaged area mean fewer opportunities to purchase fresh fruit and vegetables in the area? Findings from the Brisbane food study. Health & Place. 2006;12(3):306-319.
19. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: A systematic review. Health & place. 2012;18(5):1172-1187.
20. Statistics Canada. Taking a Census of Population. 2006; http://www12.statcan.ca/census-recensement/2006/ref/dict/overview-apercu/pop2-eng.cfm#06. Accessed 22-02-2016.

21. Beaulac J, Kristjansson E, Cummins S. A systematic review of food deserts, 1966-2007. Preventing Chronic Disease. 2009;6(3):A105.

22. Ren S, Lai H, Tong W, Aminzadeh M, Hou X, Lai S. Nonparametric bootstrapping for hierarchical data. Journal of Applied Statistics. 2010;37(9).

23. Jiao J, Moudon AV, Ulmer J, Hurvitz PM, Drewnowski A. How to identify food deserts: Measuring physical and economic access to supermarkets in King County, Washington. American Journal of Public Health. 2012;102(10):e32-e39.

24. Zenk SN, Schulz AJ, Matthews SA, et al. Activity space environment and dietary and physical activity behaviors: a pilot study. Health & Place. Sep 2011;17(5):1150-1161.

25. Drewnowski A, Aggarwal A, Hurvitz PM, Monsivais P, Moudon AV. Obesity and supermarket access: proximity or price? American Journal of Public Health. 2012;102(8):e74-e80.

26. Mackenbach JD, Burgoine T, Lakerveld J, et al. Accessibility and affordability of supermarkets: associations with the dash diet. Am J Prev Med. 2017;53(1):55-62.

27. Wan X, Evers PT, Dresner ME. Too much of a good thing: The impact of product variety on operations and sales performance. Journal of Operations Management. 2012;30(4):316-324.

28. Andreyeva T, Blumenthal DM, Schwartz MB, Long MW, Brownell KD. Availability and prices of foods across stores and neighborhoods: the case of New Haven, Connecticut. Health Aff (Millwood). Sep-Oct 2008;27(5):1381-1388.

29. Byker Shanks C, Ahmed S, Smith T, et al. Availability, price, and quality of fruits and vegetables in 12 rural Montana counties, 2014. Preventing Chronic Disease. 2015;12:E128.

30. Black C, Moon G, Baird J. Dietary inequalities: What is the evidence for the effect of the neighbourhood food environment? Health & Place. 2014;27:229-242.

31. Polsky JY, Moineddin R, Glazier RH, Dunn JR, Booth GL. Foodscapes of southern Ontario: neighbourhood deprivation and access to healthy and unhealthy food retail. Canadian Journal of Public Health-Revue Canadienne De Sante Publique. Sep-Oct 2014;105(5):e369-375.

32. Wang J, Engler-Stringer R, Muhajarine N. Assessing the consumer food environment in restaurants by neighbourhood distress level across Saskatoon, Saskatchewan. Canadian Journal of Dietetic Practice and Research. 2015;77(1):9-16.

33. Gould AC, Apparicio P, Cloutier MS. Classifying Neighbourhoods by Level of Access to Stores Selling Fresh Fruit and Vegetables and Groceries: Identifying Problematic Areas in the City of Gatineau, Quebec. Canadian Journal of Public Health-Revue Canadienne De Sante Publique. Nov-Dec 2012;103(6):E433-E437.

34. Black JL, Carpiano RM, Fleming S, Lauster N. Exploring the distribution of food stores in British Columbia: associations with neighbourhood socio-demographic factors and urban form. Health & Place. 2011;17(4):961-970.

35. Cannuscio CC, Tappe K, Hillier A, Buttenheim A, Karpyn A, Glanz K. Urban food environments and residents' shopping behaviors. Am J Prev Med. 2013;45(5):606-614.

36. Leone AF, Rigby S, Betterley C, et al. Store type and demographic influence on the availability and price of healthful foods, Leon County, Florida, 2008. Preventing Chronic Disease. Nov 2011;8(6):A140.

37. Sharkey JD, WE; Nalty, C,. Convenience stores and the marketing of foods and beverages through product assortment. Am J Prev Med. 2012;43(3):S109-S115.

38. Charreire H, Casey R, Salze P, et al. Measuring the food environment using geographical information systems: a methodological review. Public Health Nutrition. 2010;13(11):1773-1785.

5.5 Supplemental Tables and Figures

Figure 5.3: Flow chart

*In the standard basket

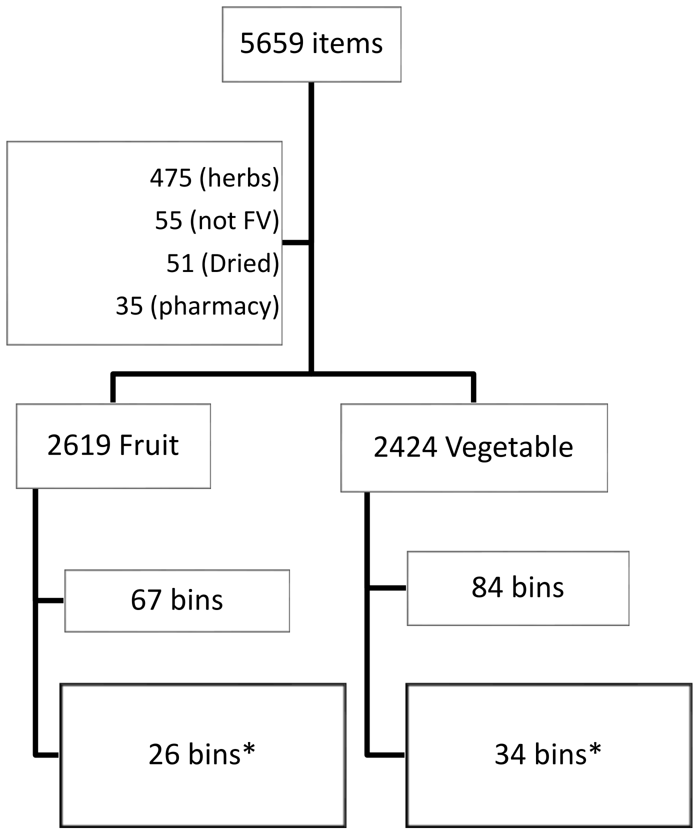


Figure 5.4: Fruit and vegetable standard basket

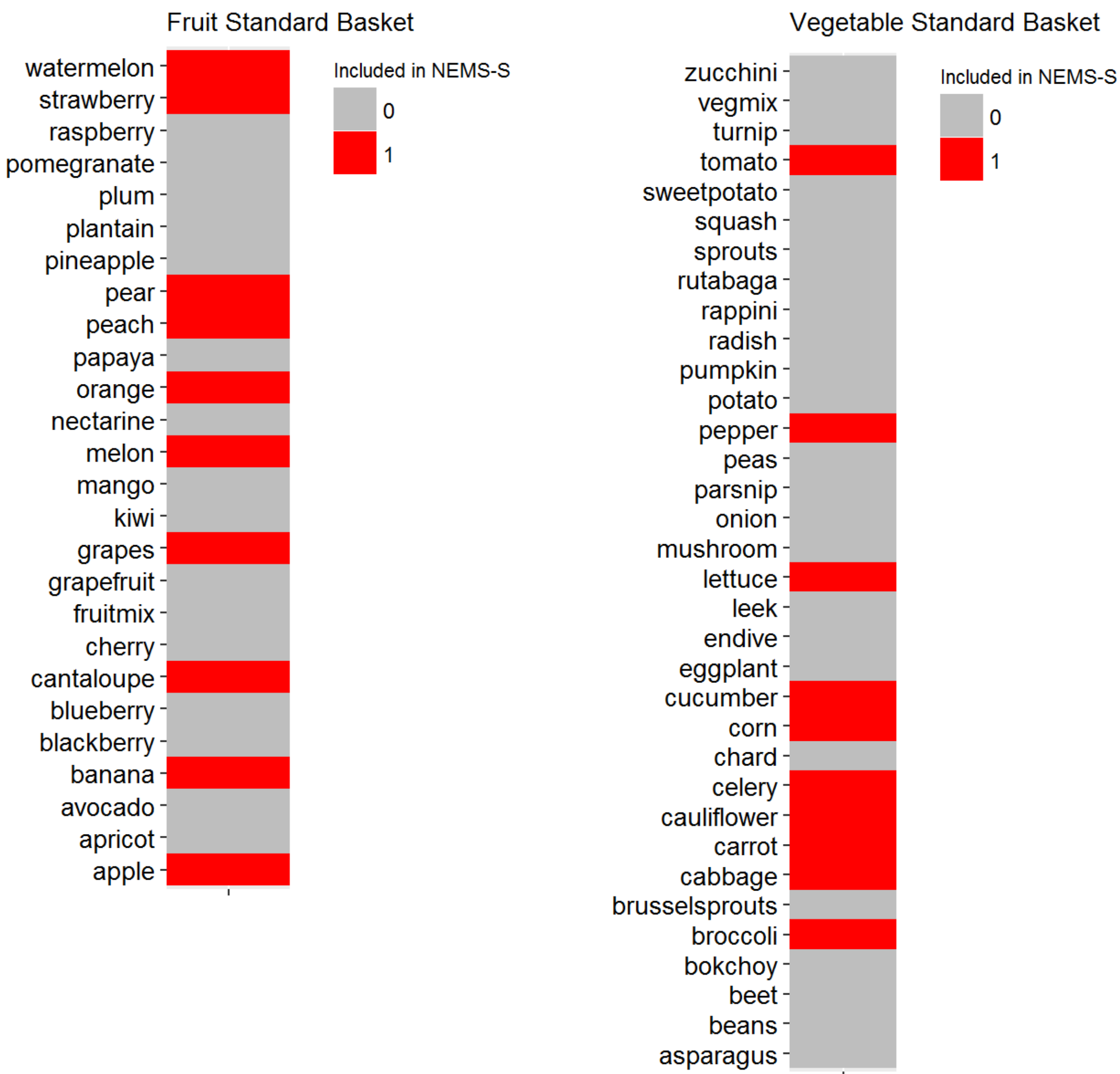


Table 5.4: Availability and variety by income level

Median Income[i]	Available (SD)[ii]	Total items (SD)	Items per bin (SD)[iii]	Stores
High (> $60,000)	0.94 (0.23)	310.08 (74.66)	5.51 (5.83)	78
Low (< 60,000)	0.91 (0.28)	295.94 (77.82)	5.31 (5.69)	90

i. Median household income of regions
ii. Average proportion of the standard basket available stores in the respective income stratum
iii. Number of 'versions' of each fruit and vegetable product (bin), i.e. variety measure

Table 5.5: Marginal predications for variety (number of items per bin)

Chain price	Prediction	95% Confidence Interval

High	7.04	6.84	7.24
Medium	6.27	6.08	6.47
Low	5.78	5.67	5.93
Income Level			
High Income	6.43	6.26	6.6
Low Income	6.29	6.12	6.47

Table 5.6: All results for availability

Outcome: Availability (Standard Basket); Linear model			
	Coefficient*	**95% Confidence Interval**	
(Intercept)	0.76	0.703	0.818
levelmed	-0.056	-0.085	-0.026
levellow	-0.072	-0.106	-0.037
income_cutlow.income	-0.021	-0.048	0.006
bs(year.month, 4)1	0.025	-0.047	0.097
bs(year.month, 4)2	0.176	0.106	0.247
bs(year.month, 4)3	0.124	0.069	0.18
bs(year.month, 4)4	0.141	0.087	0.195
log(pop.density)	0.027	0.02	0.034
levelmed:income_cutlow.income	0.025	-0.018	0.068
levellow:income_cutlow.income	-0.005	-0.058	0.048
*Represents proportion difference from linear model with Income level, chain price point and interaction term adjusted for 4-knot month-year index, with robust standard errors			

Table 5.7: All results for variety

Outcome: Variety (Items per bin); Quasipoisson model			
	Coefficient*	**95% Confidence Interval**	
(Intercept)	1.765	1.709	1.821
levelmed	-0.135	-0.187	-0.083
levellow	-0.211	-0.241	-0.181
income_cutlow.income	-0.04	-0.072	-0.008
bs(year.month, 4)1	0.028	-0.042	0.098
bs(year.month, 4)2	-0.07	-0.151	0.01
bs(year.month, 4)3	0.172	0.107	0.237
bs(year.month, 4)4	0.069	0.008	0.131
log(pop.density)	0.036	0.03	0.043
levelmed:income_cutlow.income	0.035	-0.03	0.1
levellow:income_cutlow.income	0.022	-0.019	0.064
*Represents proportion difference from linear model with Income level, chain price point and interaction term adjusted for 4-knot month-year index, with robust standard errors			

Chapter 6

Study Three: Temporal and seasonal variation in the affordability of fruits and vegetables

In this objective, I measure the seasonal amplitude and temporal trend of fruit and vegetable prices, and determine whether these vary by overall regional affordability. Area and seasonal price differences can create barriers related to affordability when shopping in certain areas or higher price seasons. I use a method used previously to study seasonal influenza. In this method, a truncated Fourier series is used in a harmonic regression model to estimate the association between time cycles against, in my case, fruit and vegetable prices. The amplitude of the resulting wave can be thought of as a measure of the seasonal fluctuation in price.

The amplitude can be derived indirectly from the coefficients estimated in the harmonic regression model. However, its uncertainty, or standard error, cannot come from the model. To address this problem, I used a bootstrap to derive confidence intervals for the amplitude. Simple bootstrap, where data are randomly sampled with replacement, is not viable because of the autocorrelation between price observations. In hierarchical bootstrap, time clusters are randomly sampled with replacement, and data within specific time clusters are sampled without replacement. The 97.5th and 2.5th percentile of the distribution of estimates from each bootstrapped dataset determines the 95% confidence interval of the estimate.

This work addresses the need for studies that incorporate a temporal dimension in food environments research, which has remained an important gap to date. This evidence can help to provide a more comprehensive understanding of the food environment, and suggest the possibility for additional avenues for public health interventions related to seasons and time.

6.1 Abstract

Objective

Affordability is an important barrier to consuming more fruits and vegetables. Food environments research measuring affordability has largely focused on area-level variation, with limited work considering variation seasonally or over time. The nature and magnitude of this variation is thus unclear, despite its potential to create periods where healthy foods become unaffordable.

Methods

We used weekly, automatically-collected grocery store scanner data (2008-2013) in Quebec, Canada to measure regional temporal variation in fruit and vegetable prices. We used the average weekly price of a standard basket of fruits and vegetables to classify regions into high and low price clusters. We then used a harmonic linear regression model of price with a linear and quadratic time trend, to derive the yearly change and seasonal amplitude. This was done for the standard basket and individual fruit and vegetable products. Using hierarchical bootstrapping at 3-week blocks, we derived a confidence interval for the seasonal amplitude.

Results

The mean weekly price for five daily servings of fruits and vegetables was $19 (SD=1.84), which was consistent over time. The average price was $2.37 (95% CI 2.3, 2.43) higher in high price regions. The average seasonal amplitude was $1.29 (95% CI 1.12, 1.48) cents. High-price regions were $1.33 (95% CI 1.16, 1.51) more per week in the high season versus the low season, compared to $1.20 (95% CI 1.03, 1.39) in low-price regions. Fruits had larger seasonal variation in price than vegetables.

Conclusion

Seasonal price variations are substantial. This variation potentially poses a barrier to fruit and vegetable consumption, especially in less affordable areas. Awareness campaigns and season-specific mitigation strategies may be appropriate public health interventions.

6.2 Main Text

6.2.1 Background

Fruits and vegetables are an important component of a healthy diet, which can prevent chronic disease and illness. The food environment can influence a person's diet by creating differential opportunities to eat healthy and unhealthy foods. Affordability, one dimension of the food environment, is the most commonly cited barrier to eating more fruits and vegetables, particularly among low income groups.[1-4] Although there is some contradictory evidence,[5] healthy diets are generally more expensive than less-healthy options in high income countries.[6-14] Lower prices are associated with less BMI gain, particularly among lower income children,[15] and each 10 cent increase in price per serving for fruits and vegetables decreased the odds of having the item at home by 23% in the U.S.[16]

Affordability of healthy foods varies regionally, though not necessarily according to regions' socio-economic characteristics. Interest in regional variation stems from *deprivation amplification theory,* which suggests that the characteristics of deprived areas amplify the effects of individual disadvantage.[17] In the U.S., prices are sometimes more expensive in ethnic-minority inner-cities than in suburbs,[18,19] though there is little other evidence including in Canada[20] that affordability varies systematically as such.[21-24] However, supermarkets still have different prices based on other regional characteristics such as operating costs and local demand elasticity.[18] As people with lower income require a greater proportion of their income to purchase the same item,[25] fruits and vegetables can still be unaffordable or perceived as unaffordable in higher versus lower price regions. Spatial variation in affordability thus remains an important subject of food environments research.

While studies have explored *where* affordability is less, few have measured *when* it may be more problematic. Fruits and vegetables are among the most volatile components of the consumer price index,[26] and their prices increased faster than for other food groups in the 2000s.[13] Prices are cheaper overall in harvest seasons,[10,27,28] though the difference depends on the product.[27] Fruit prices were previously found to fluctuate by about CAD 10 cents per serving in Montreal, Quebec within a year.[28] Such measures have rarely been estimated. Seasonal differences can create acute periods of lesser affordability and may depend on regions' average prices. For example, rural areas are sometimes found to be more expensive,[24] but it is unclear whether they also experience greater seasonal fluctuation that could compound any background lack of affordability. Poorer households are particularly susceptible to higher regional prices, seasonal fluctuations and temporal increases.

In this study, we aim to measure temporal and seasonal disparities in fruit and vegetable affordability and determine if these disparities depend on average regional prices. Achieving this aim will help to further characterize economic food access and provide new insight into the barriers to consuming healthy foods.

6.2.2 **Methods**

Variables

Price data were obtained from a large marketing firm that consolidates automatically-collected digital sales records from cash registers in retail stores. Data were collected on a weekly basis from January 2008 to December 2013 (inclusive) in Quebec, Canada. The store sample was selected based on a cluster-stratified random sample and includes supermarkets with more than $2 million in annual revenue. Frozen fruit and vegetable products and canned products were excluded from the sample. We additionally excluded store-brand products, which had limited data available to protect chain confidentiality, organic products, because organic food consumers have particular characteristics,[29] and dried foods, which have different nutritional value.[30]

A region was defined using *réseaux locaux de services* (RLS) boundaries. These are administrative areas for health and social service delivery. We restricted the analysis to regions containing sampled stores so that we would have direct measurements of food price in each region (n = 57 out of 96 total RLS). Regions with missing price data in more than 10% of weeks were also excluded. Information on the material deprivation index[31] and degree of rurality were obtained from the 2006 Canadian Census and 2011 National Household Survey.[32] A generalized linear model was used to extrapolate rurality and the material deprivation index to each year between the two measured points.

Affordability was measured as the average weekly price to meet the recommended five servings of a standard basket of commonly consumed fruits and vegetables. The basket consisted of 29 vegetable and 23 fruit products which ranked in the top half of sales volume and dollars, and were sold in at least 90% of stores in 2008. The standard basket price was the average price per 100 or 140 gram serving of each vegetable and fruit product, respectively, times 35 for five servings per day for seven days. We truncated the most and least expensive items in each product category (5^{th} and 95^{th} percentile), as is common practice.[33] Prices were standardized to the 2010 consumer price index for fruits and vegetables. In sensitivity analyses, we weighted the weekly price by the average yearly market share of the fruit and vegetable products to reflect consumer demand (Section 6.3.3).

Analysis

Classifying regional affordability

We classified regions based on the average monthly price of the standard basket in stores in that region using k-means clustering. K-means is an unsupervised machine learning algorithm useful to discover patterns in data. The set of regional time series were clustered according to their similarity in absolute values and the shape of the series.[34,35] Similarity was defined as Euclidian distance. We employed a 2-cluster solution to classify regions into 'high' (1) and 'low' (0) price. To determine whether these regions differed based on material deprivation (dichotomized as 'deprived' at quintile >3) and degree of rurality (dichotomized as 'rural' at zone >3 using Census classifications), we used a linear regression model of these predictors against the price cluster. We used robust standard errors clustered at the region level.

Price information for regions with less than 10% missing data was imputed to complete the clustering analysis. Imputed values were derived by smoothing the time series based on an autoregressive integrated moving average (ARIMA) model. The best model was selected through the Akaike and Bayesian Information Criterions (AIC, BIC) (Section 6.3.1).

Temporal and seasonal trends

We assessed linear and quadratic temporal trends using a regression model of week against the mean price. Seasonal fluctuations were first measured empirically, and then quantified by estimating the amplitude of a harmonic regression model. We averaged monthly prices over the six years to determine the empirical trend, for example, January 2008-2013, February 2008-2013, etc. This was done overall and by regional affordability level.

To measure seasonal amplitude, we borrowed from an approach previously used to study seasonal influenza patterns.[36] We modelled our data using a harmonic regression model based on a 52.17 week annual period (Equation 6-1). This model is based on a truncated Fourier series, which expresses a time series as a sum of cosine functions. We determined whether including additional harmonics and the time trend improved model fit using AIC and BIC. Using the best model, we extracted the total amplitude from the coefficients using Equation 6-2. The amplitude is the height of the peaks relative to the mean, and thus measures the seasonal fluctuation.

Equation 6-1: Regression model

$$M_t = a_0 + a_1 \cos\frac{2\pi t}{T} + a_2 \sin\frac{2\pi t}{T} + \varepsilon_t$$

M_t = mean price at week t
t = running index for time (week)
T = period (52.17 weeks)
ε_t = normally distributed errors

Equation 6-2: Amplitude

$$A = \sqrt{{a_1}^2 + {a_2}^2}$$

a_1 and a_2 estimated from Equation 6-1

The model was run overall, and stratified by regional affordability level. We derived confidence intervals for the amplitude using a hierarchical bootstrap with 250 replicates clustered at three-week blocks to account for residual autocorrelation between the price observations.[37] The resulting confidence intervals provided a typical measure of imprecision around our estimate. Amplitudes were additionally estimated for individual fruit and vegetable products to help explain the observed trends.

All main analyses were done in *R* (R Development Core Team), though PostgreSQL was used to manage the data.

6.2.3 Results

Of the 57 regions with price information, 47 had 10% or fewer weeks with missing data and were available for analysis. We imputed time series data for 9 of these regions (see Section 6.3.1 for imputation description and results). The average price for five servings of fruits and vegetables remained relatively consistent over the time period and there was evident regional and monthly variation in price (Table 6.1).

Table 6.1: Mean weekly price (2010 CAD) for five servings of fruits and vegetables

Year	Mean Price ($)	SD (Monthly)	SD (Regional)	Stores
2008	20.27	1.26	1.75	163
2009	19.07	0.75	1.62	150
2010	19.25	0.74	1.84	148
2011	19.18	0.57	2.02	155
2012	19.38	0.51	1.97	153
2013	19.22	0.73	1.8	153

Based on the k-means clustering, the weekly fruit and vegetable price was $2.37 (95% CI =2.3, 2.43) higher in the high-price cluster. There were fewer regions in the low-price cluster, though the clusters were similar on other characteristics (Table 6.2). There were no evident differences in the proportion of regions classified as deprived (average difference =0.08 [95% CI 0.33, -0.17]), or rural (average difference = 0.15 [95% CI 0.4, -0.11]).

Table 6.2: Regional price clustering results

Price Cluster	Regions	Stores per Region (SD)	Regional population in 1000s (SD)	Chains per Region (SD)	Mean Price $ (SD)
Low	13	2.65 (1.60)	97.6 (70)	2.35 (1.26)	17.51 (1.71)
High	33	3.37 (2.61)	139.4 (88)	2.85 (1.79)	19.87 (1.61)

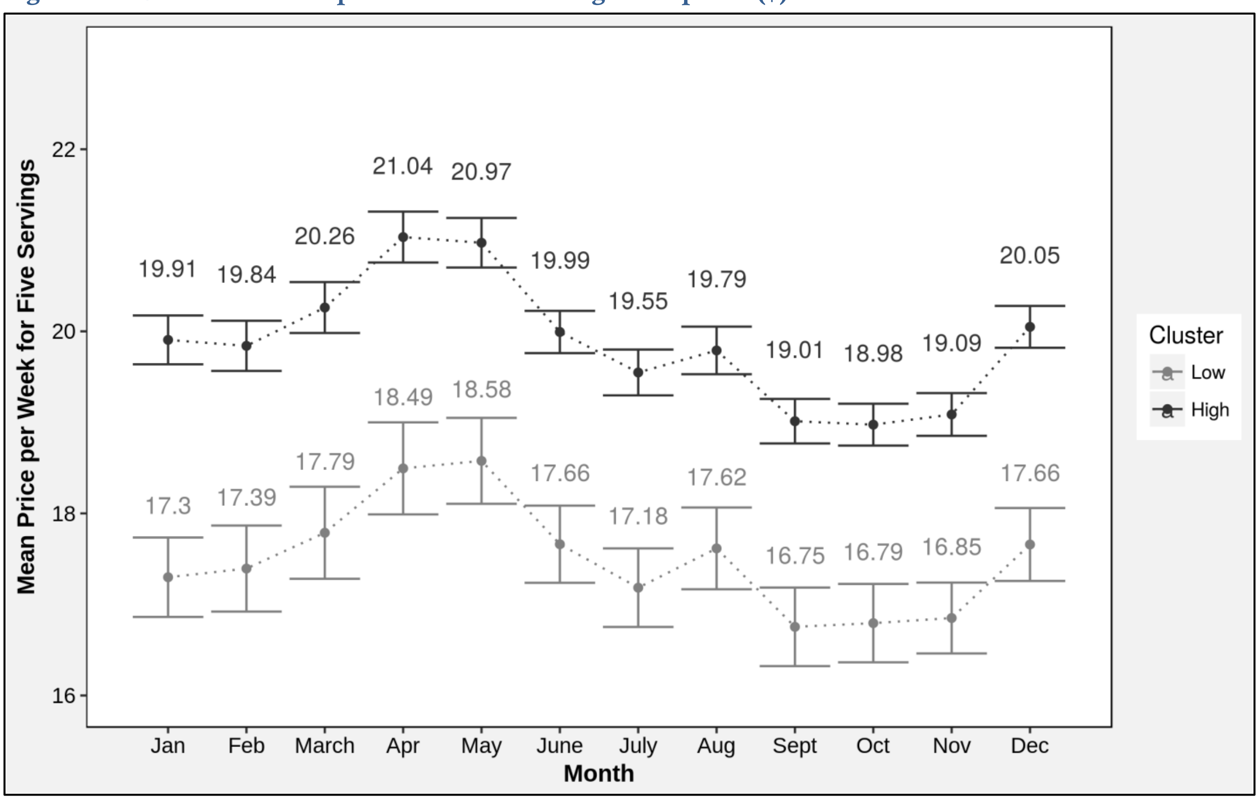
Fruit and vegetable prices decreased slightly over time. The decrease in the low-price cluster was faster than the decrease in the high price cluster (Table 6.3). The quadratic term for the model was close to zero, though it did improve model fit ($p < 0.001$ for likelihood ratio test) and was thus retained in the harmonic regression model below.

Table 6.3: Time trend results from linear model of weekly index on mean weekly price

Price cluster	Intercept	Time (Mean weekly change)
overall	20.39 (20.13; 20.65)	-0.02 (-0.02;-0.01)
high	20.89 (20.62; 21.16)	-0.01 (-0.02; -0.01)
low	19.11 (18.86; 19.35)	-0.02 (-0.03; -0.02)

Seasonal variation in affordability was evident (Figure 6.1). Overall, the average price per week for five servings was $18.69, but during the most expensive month, the average was $21.04. During the peak month, five servings of fruits and vegetables thus cost $2.35 more per week than the average, and $4.28 more than in the minimum month. The difference between the average and peak price was $0.09 cents higher in high price regions than low price regions ($1.07 versus $1.16).

Figure 6.1: Overall seasonal pattern in fruit and vegetable prices ($)



The best harmonic regression model included terms for the first three harmonics as well as the linear and quadratic time trend (Figure 6.2). The full model selection is availability in Section 6.3.1.

The average amplitude, or maximum difference from the overall average price, was \$1.29 (95% CI 1.12, 1.48). High price regions were on average \$1.33 (95% CI 1.16, 1.51) more per week to meet five servings of fruits and vegetables in the high season. This peak was higher than in low price regions, where the difference was \$1.20 (95% CI 1.03, 1.39).(Table 4.4) Weighting by market shares did not alter our conclusions. The average temporal trend and amplitude estimates were slightly lower in magnitude, but not substantively different. The seasonal pattern was similar (Section 6.3.3).

Figure 6.2: Seasonal pattern estimated using regression model with first three harmonics

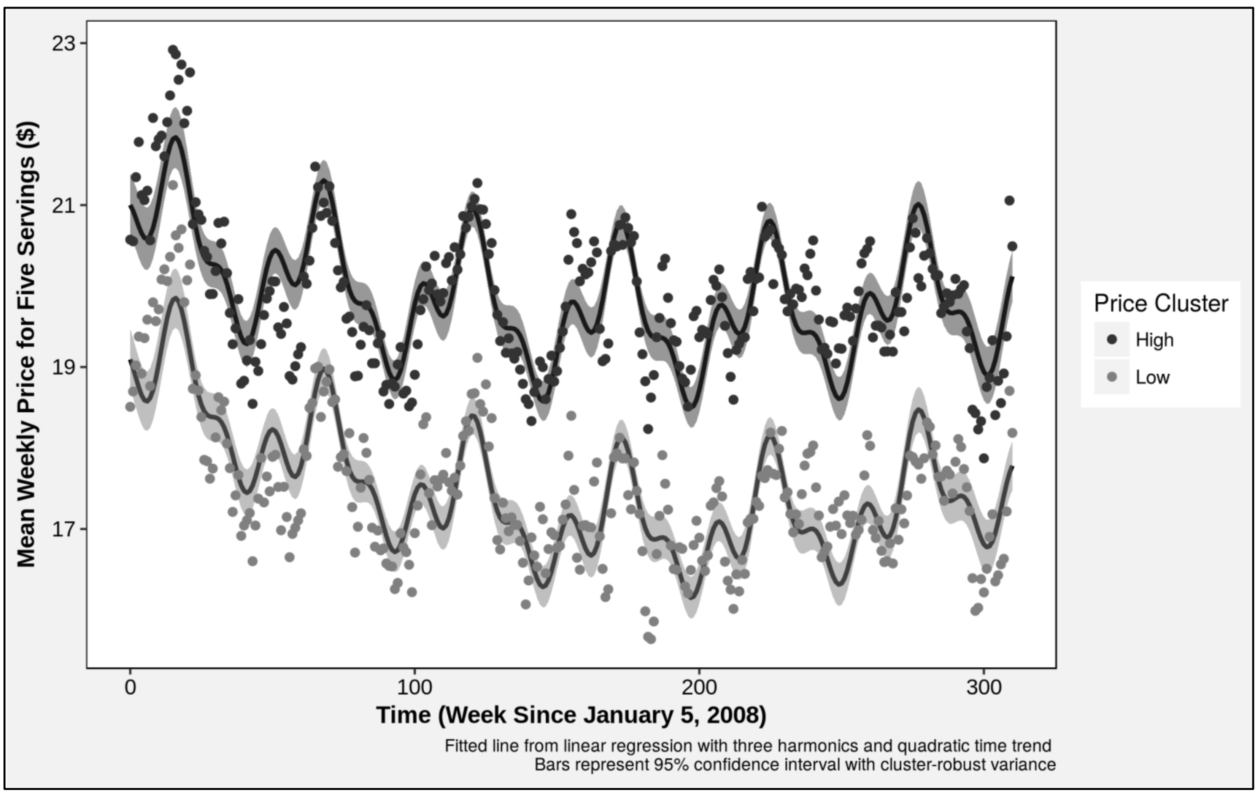


Table 6.4: Estimated amplitudes from regression model in dollar difference from overall mean

Price cluster	Amplitude[1]
Overall	1.29 (95% CI 1.12;1.48)[2]
High	1.33 (95% CI 1.16;1.51)
Low	1.20 (95% CI (1.03;1.39)

1. Difference between average yearly price and price in peak season

2. Model was adjusted for a linear and quadratic time trend and confidence intervals were derived from cluster bootstrap using 3-week blocks

Individual products' relative seasonal variation varied considerably. The amplitudes for fruits were higher than for vegetables and products' seasonal variation also differed between the regional affordability clusters (Figure 6.3 and 6.4). Generally, fruits varied more than vegetables. The five vegetable products that had the greatest difference between the two regional price clusters were cabbage, brussel sprouts, beans, eggplant, and peas (average price differences of 4.15, 1.37, 1.19, 1.02, and 1, respectively for 35 servings), while for fruits it was pomegranate, raspberry, cherry, banana and apricot (average differences of 5.21, 4.04, 3.36, 1.65, and 1.24, respectively for 35 servings) (See Section 6.5 for full results). Notably, only stores that contain the product in each month contribute to the price estimate for that month, which is seasonally variable.

Figure 6.3: Fruit seasonal amplitudes by regional price cluster

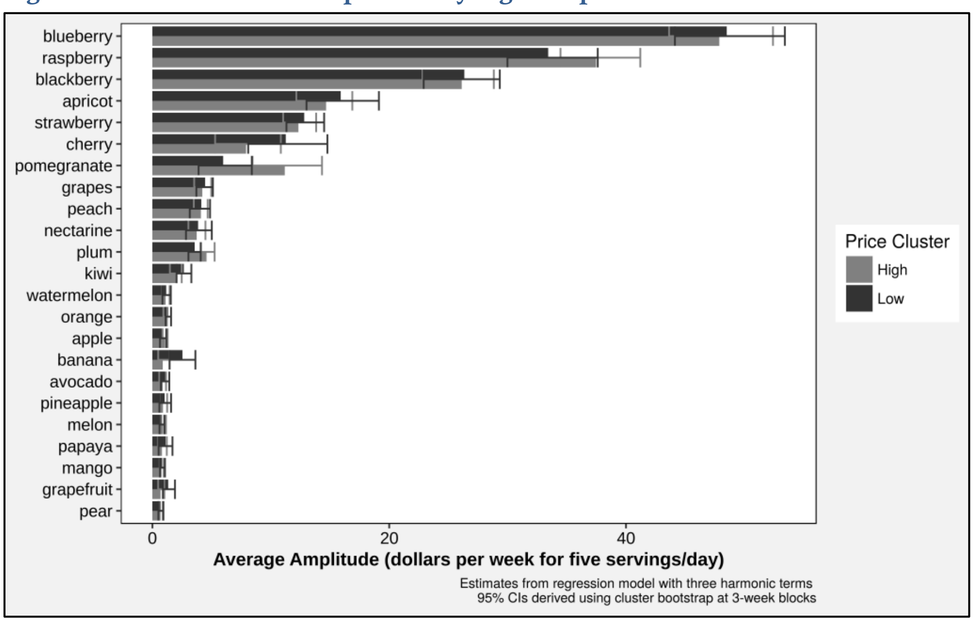
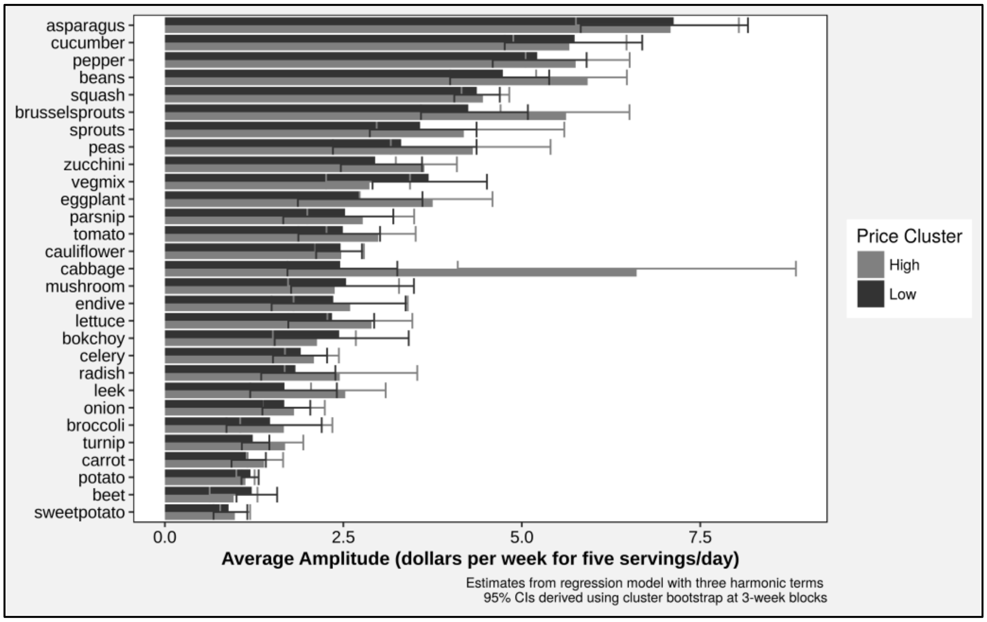


Figure 6.4: Vegetable seasonal amplitudes by regional price cluster



6.2.4 **Discussion**

Fruit and vegetable affordability varies substantially depending on the season and region, though we did not find a longer-term time trend. The regional and season price disparities could make fruit and vegetable purchasing out of reach in more expensive areas, or limit the opportunity to have a variety of products during more expensive seasons. Strategies that control for seasonality[38,39] can mask this important variation. Understanding seasonal food environment affordability can help design appropriate interventions to support healthier purchasing. Further, the differences represent an additional source of inequality in opportunities to eat healthy.

Awareness campaigns usually aim to emphasize the importance of consuming fruits and vegetables, but seasonality could represent an additional component to such campaigns. Stocking products in low price seasons and freezing could translate to significant savings in high price seasons.[40] While canned produce prices are more consistent seasonally, the taste of fresh produce is preferred,[41] making this option less attractive. Additionally, seasonality is product-specific. Knowing which products are particularly susceptible to seasonality can support consumers to seek seasonally-cheaper products. We find berries and asparagus, for instance, varied considerably, making them particularly expensive choices out-of-season relative to in-season prices. Such strategies could potentially help to mitigate lack of affordability. To the extent that variety is important to purchasing decisions, however, increased awareness of seasonality alone cannot eliminate susceptibility to static and dynamic food environment

The regional and seasonal differences could further disparities in economic access to fruits and vegetables for lower income households. Overall, we found that shopping in high price areas cost $2.37 (95% CI =2.3, 2.43) more per person per week, or approximately $10 per week for a family of four, to meet the recommended amount of fruits and vegetables. High price regions are not necessarily more likely to be deprived,[42] however such a difference can still inhibit consumption for low income households located in high price regions. These barriers may be compounded by season. While most products' seasonal variation was similar in different regions, bananas, for example, had relatively higher seasonal fluctuation in high price regions. Previous work has found that seasonality is a concern for rural and/or low income households.[41] For instance, they report limiting their purchasing depending on season.[43] It is possible that chains catering to different budget levels also vary differentially by season and over time, potentially due to micromarketing according to consumers' seasonal expectations and the elasticity of demand.[44]

Limitations

The main limitation of this study is that we only included major chains. In the fall and summer, there may be other options to purchase produce such as farmers' markets.[45] Where these options are cheaper than supermarkets, the average regional seasonal price difference could shrink. However, the price difference between farmers' markets and supermarkets depends on the product and, out-of-season, supermarkets tend to be cheaper.[27] There is also the remaining potential for correlation between observations, despite our efforts to mitigate this using the clustered bootstrap. However, given our estimates were considerably higher than the null value of 0, it is unlikely that the remaining auto-correlation would result in true confidence intervals that end up including the null value and thus alter our conclusion.

6.2.5 **Conclusion**

Seasonal and regional price disparities in fruits and vegetables are substantial. Compounding lower affordability in high price regions, seasonal variation could potentially render fruits and vegetables unaffordable at certain times. Awareness campaigns and season-specific mitigation strategies could represent another avenue for public health interventions. Future research could capitalize the vast amounts of data available through digital grocery store scanner records to investigate the dynamic nature of the food environment.

6.3 Additional Analyses

6.3.1 Modelling the seasonal pattern in price time series

The primary analysis used three harmonics of a Fourier series (Equation 6-3).

Equation 6-3: Primary regression model

$$M_t = a_0 + a_1 \cos\frac{2\pi t}{T} + a_2 \sin\frac{2\pi t}{T} + \beta_1 \cos\frac{4\pi t}{T} + \beta_2 \sin\frac{4\pi t}{T} + \chi_1 \cos\frac{6\pi t}{T} + \chi_2 \sin\frac{6\pi t}{T} + t + t^2 + \varepsilon_t$$

T in our case was one year, or 52.17 weeks, because the seasonal patterns were weekly within a year. t was a running time index indicating the week and y was the weekly mean price for five servings.

I compared the fit of models with one, two, three and four harmonics using the AIC and BIC, as well as the adjusted R-squared. In addition, the price values predicted from each model were plotted (Figure 6.5). The models with two and four harmonics were also a good fit to estimate the seasonal trend, but the model with three harmonics was superior without over-fitting (Table 6.5).

Figure 6.5: Predicted values from models with different harmonics

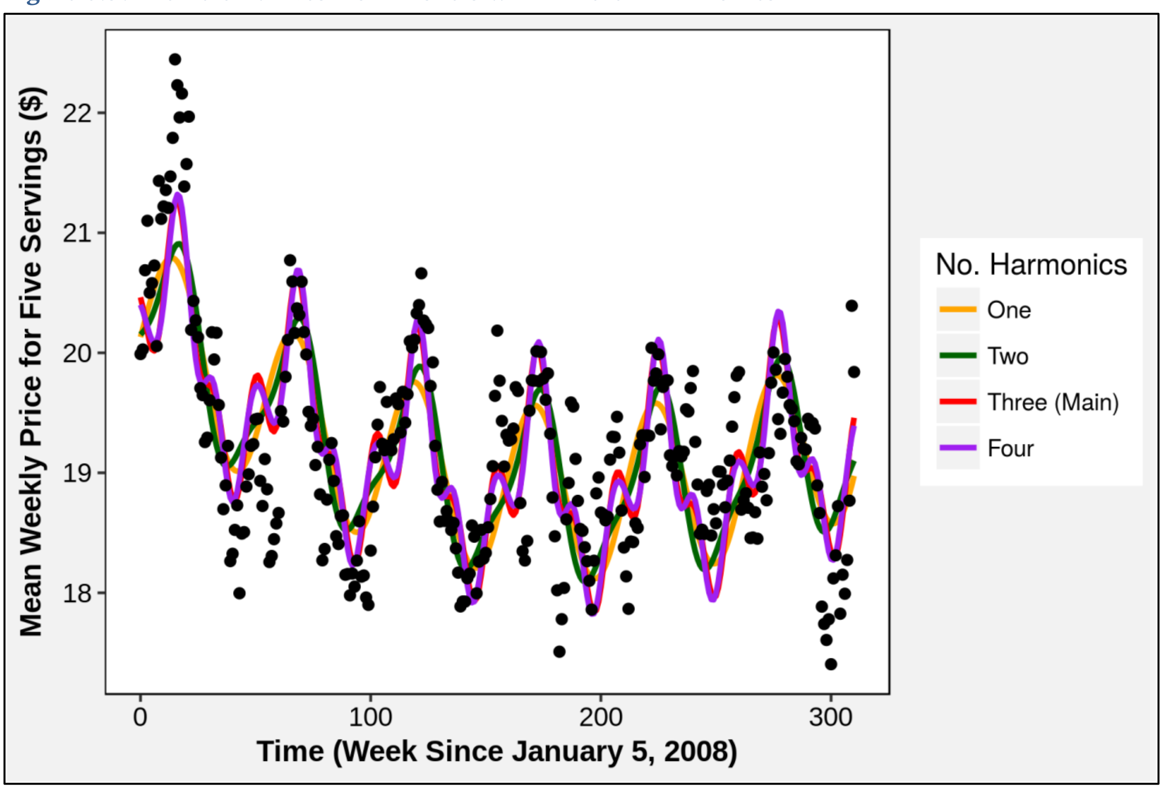


Table 6.5: Fit criteria for models with increasing complexity

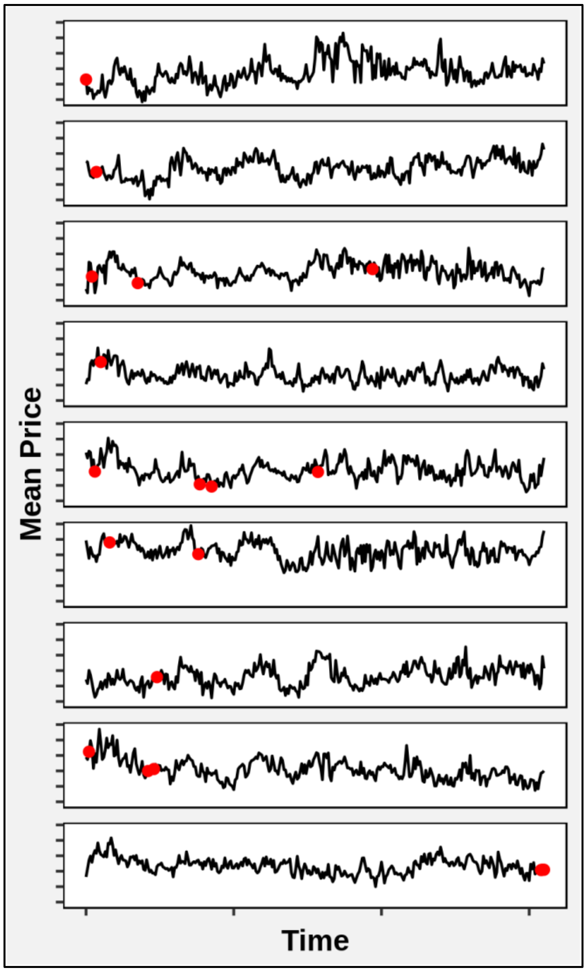
Harmonics	AIC	BIC	Adj.R^2
1	568.16	590.60	0.55
2	552.76	582.68	0.57
3	**485.31**	**522.71**	**0.66**
4	485.61	530.49	0.66

6.3.2 Imputation description and results

We imputed missing points for nine regions with less than 10% of missing data. This was done by modelling the overall time series process and then using the results to smooth over the missing points. The process was modeled using an ARIMA model. ARIMA models are often used to model time series based on the correlation between nearby values and inertia. They are popular for forecasting. The auto-regressive (AR) part models the time series based on its own previous values, while the moving average (MA) effectively measures the time series' memory of 'shocks', i.e how long a large spike in the time series affects future values in the time series. Auto-regressive and moving average values for this model are tried iteratively until the best fit is achieved.

In my case, the selection was automated based on several possible models and choosing the one with the best AIC and BIC. This was repeated for each region with missing values. I allowed for the models to be non-stationary and seasonal, meaning the mean and variance could change over time. Once the best ARIMA model was chosen, I used the fitted values with a Kalman smoother to 'fill in' the missing points. This method allows the best value for each point to be estimated based on the whole time series (Figure 6.6).

Figure 6.6: Imputation results
(each panel is a region and red dots are the imputed points)



6.3.3 Weighting average prices by products' market share

Market weights reflect the purchasing patterns of a consumer group. A simple average of product prices may be skewed towards expensive products that few people purchase. Instead, a weighted average using the market weight of each product produces an overall mean that reflects what consumers like to buy.

I derived the temporal trend, seasonal trend and amplitude using the price weighted by market share. The weights were the volume (servings) sold of an individual product over the total volume of fruits and vegetables sold in year 2008. The weights were derived separately for fruits and vegetables (Equation 6-4).

For each product, p, and total servings sold, T:

Equation 6-4: Market share

$$\frac{\sum_{sales} X_p}{\sum_{sales} X_T}$$

Both the temporal and seasonal trends were similar but lower in magnitude (Table 6.6) and amplitude (Table 6.7), respectively. The peaks and troughs in seasonal prices followed a similar pattern (Figure 6.7). As expected, the market weights ultimately smoothed the patterns slightly, but did not change the conclusion.

Figure 6.7: Overall seasonal pattern with average prices weighted by product market share

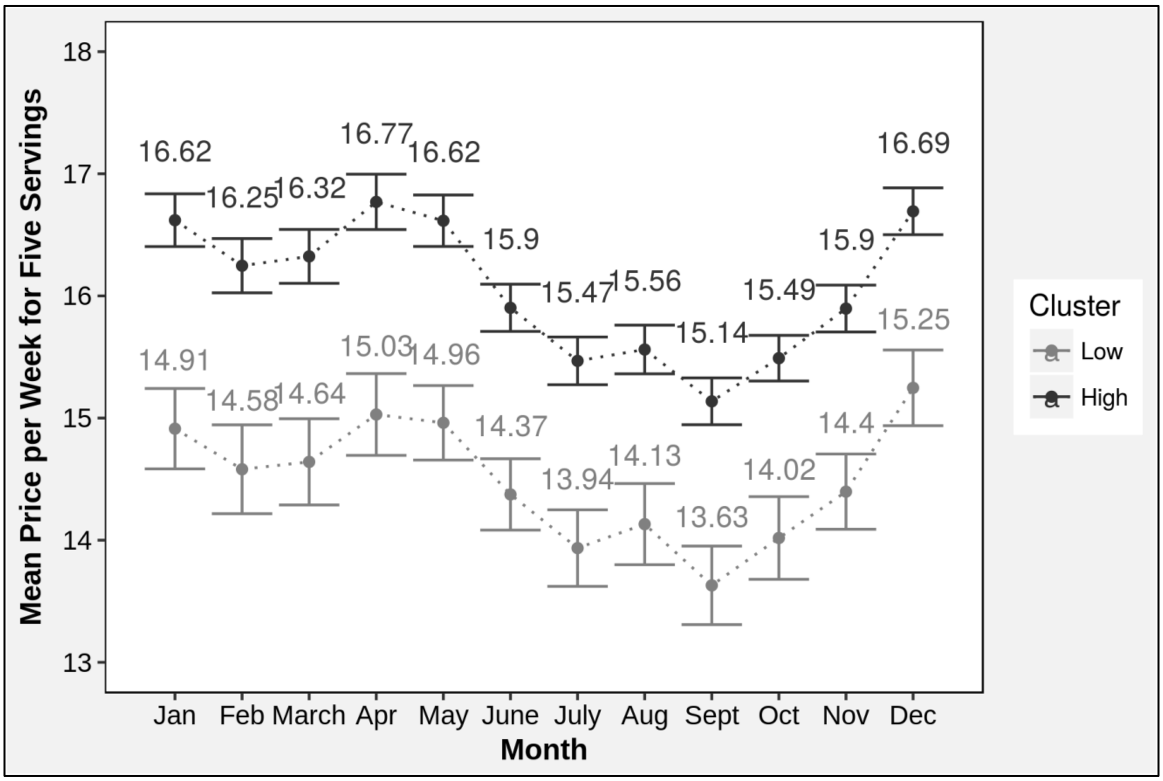


Table 6.6: Time trend coefficients in original (a) and sensitivity (b) analysis

Price cluster	a.time	b.time
Overall	-0.02 (-0.02, -0.01)	-0.01 (-0.01, -0.01)
High	-0.01 (-0.02, -0.01)	-0.01 (-0.01, -0.01)
Low	-0.02 (-0.03, -0.02)	-0.02 (-0.02, -0.01)

Table 6.7: Amplitude results with average standard basket price weighted by product market share

Price cluster	Amplitude from weighted results	Amplitude from primary analysis
Overall	1.15 (1.04, 1.25)	1.29 (1.12, 1.48)
High	1.18 (1.07, 1.28)	1.33 (1.16, 1.51)
Low	1.09 (0.96, 1.18)	1.20 (1.03, 1.39)

6.4 References

1. John JH, Ziebland S. Reported barriers to eating more fruit and vegetables before and after participation in a randomized controlled trial: a qualitative study. Health Education Research. 2004;19(2):165-174.
2. Haynes-Maslow L. A qualitative study of perceived barriers to fruit and vegetable consumption among low-income populations, North Carolina, 2011. Preventing Chronic Disease. 2013;10.
3. Dijkstra SC, Neter JE, van Stralen MM, et al. The role of perceived barriers in explaining socio-economic status differences in adherence to the fruit, vegetable and fish guidelines in older adults: a mediation study. Public Health Nutrition. 2015;18(05):797-808.
4. Ma Y, Ailawadi KL, Grewal D. Soda versus cereal and sugar versus fat: Drivers of healthful food intake and the impact of diabetes diagnosis. Journal of Marketing. 2013;77(3):101-120.
5. Black C, Moon G, Baird J. Dietary inequalities: What is the evidence for the effect of the neighbourhood food environment? Health & Place. 2014;27:229-242.
6. Hakansson A. Has it become increasingly expensive to follow a nutritious diet? Insights from a new price index for nutritious diets in Sweden 1980-2012. Food Nutr Res. 2015;59:26932.
7. Rao M, Afshin A, Singh G, Mozaffarian D. Do healthier foods and diet patterns cost more than less healthy options? A systematic review and meta-analysis. BMJ open. 2013;3(12):e004277.
8. Barratt J. The cost and availability of healthy food choices in southern Derbyshire. Journal of Human Nutrition and Dietetics. 1997;10(1):63-69.
9. Burns CM, Gibbon P, Boak R, Baudinette S, Dunbar JA. Food cost and availability in a rural setting in Australia. Rural Remote Health. Oct-Dec 2004;4(4):311.
10. Cassady D, Jetter KM, Culp J. Is price a barrier to eating more fruits and vegetables for low-income families? Journal of the American Dietetic Association. 2007;107(11):1909-1915.
11. Dietitians of Canada. The cost of eating in BC 2009: Low income British Columbians can't afford healthy food. Community Nutritionists Council of BC. 2009.
12. Harrison MS, Coyne T, Lee AJ, Leonard D. The increasing cost of the basic foods required to promote health in Queensland. Medical Journal of Australia. 2007;186(1):9.
13. Monsivais P, Drewnowski A. The rising cost of low-energy-density foods. Journal of the American Dietetic Association. 2007;107(12):2071-2076.
14. Newell FD, Williams PL, Watt CG. Is the minimum enough? Affordability of a nutritious diet for minimum wage earners in Nova Scotia (2002-2012). Can J Public Health. May-Jun 2014;105(3):e158-165.
15. Sturm R, Datar A. Body mass index in elementary school children, metropolitan area food prices and food outlet density. Public Health. 2005;119(12):1059-1068.
16. Ard JD, Fitzpatrick S, Desmond RA, et al. The impact of cost on the availability of fruits and vegetables in the homes of schoolchildren in Birmingham, Alabama. American Journal of Public Health. Feb 2007;97(2):367-372.
17. Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? International Journal of Behavioral Nutrition and Physical Activity. 2007;4(1):32.
18. MacDonald JM, Nelson Jr PE. Do the poor still pay more? Food price variations in large metropolitan areas. Journal of Urban Economics. 1991;30(3):344-359.
19. Chung C, Myers SL. Do the poor pay more for food? An analysis of grocery store availability and food price disparities. Journal of Consumer Affairs. 1999;33(2):276-296.
20. Latham J, Moffat T. Determinants of variation in food cost and availability in two socioeconomically contrasting neighbourhoods of Hamilton, Ontario, Canada. Health & Place. Mar 2007;13(1):273-287.
21. Gustafson A, Hankins S, Jilcott S. Measures of the consumer food store environment: A systematic review of the evidence 2000–2011. Journal of Community Health. 2012;37(4):897-911.

22. Andreyeva T, Blumenthal DM, Schwartz MB, Long MW, Brownell KD. Availability and prices of foods across stores and neighborhoods: the case of New Haven, Connecticut. Health Aff (Millwood). Sep-Oct 2008;27(5):1381-1388.
23. Ball K, Timperio A, Crawford D. Neighbourhood socioeconomic inequalities in food access and affordability. Health & Place. 2009;15(2):578-585.
24. Hardin-Fanning F, Rayens MK. Food cost disparities in rural communities. Health Promot Pract. May 2015;16(3):383-391.
25. Darmon N, Drewnowski A. Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: a systematic review and analysis. Nutrition Reviews. 2015;73(10):643-660.
26. Statistics Canada. Table 326-0020 Consumer Price Index, monthly (2002=100).2008-2013.
27. Valpiani N, Wilde P, Rogers B, Stewart H. Patterns of fruit and vegetable availability and price competitiveness across four seasons are different in local food outlets and supermarkets. Public Health Nutrition. Oct 2015;18(15):2846-2854.
28. Ma Y, Labban A, Cherian M, Shaban-Nejad A, Buckeridge DL, Dubé L. System of indicators for the nutritional quality of marketing and food environment: product quality, availability, affordability, and promotion. Diet Quality; 2013:383-396.
29. Zepeda L, Li J. Characteristics of organic food shoppers. Journal of Agricultural and Applied Economics. 2007;39(01):17-28.
30. Lutz M, Hernández J, Henríquez C. Phenolic content and antioxidant capacity in fresh and dry fruits and vegetables grown in Chile. CyTA-Journal of Food. 2015;13(4):541-547.
31. Pampalon R, Raymond G. A deprivation index for health and welfare planning in Quebec. Chronic Diseases and Injuries in Canada. 2000;21(3):104.
32. Statistics Canada. Taking a Census of Population. 2006; http://www12.statcan.ca/census-recensement/2006/ref/dict/overview-apercu/pop2-eng.cfm#06. Accessed 22-02-2016.
33. Čížek P. General trimmed estimation: robust approach to nonlinear and limited dependent variable models. Econometric Theory. 2008;24(6):1500-1529.
34. Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T. Time-series clustering – A decade review. Information Systems. 2015;53:16-38.
35. Genolini C, Falissard B. KmL: A package to cluster longitudinal data. Computer Methods and Programs in Biomedicine. 2011;104(3):e112-e121.
36. Yu H, Alonso WJ, Feng L, et al. Characterization of regional influenza seasonality patterns in china and implications for vaccination strategies: spatio-temporal modeling of surveillance data. PLoS Medicine. 2013;10(11):e1001552.
37. Ren S, Lai H, Tong W, Aminzadeh M, Hou X, Lai S. Nonparametric bootstrapping for hierarchical data. Journal of Applied Statistics. 2010;37(9).
38. Lewis M, Lee A. Costing 'healthy' food baskets in Australia – a systematic review of food price and affordability monitoring tools, protocols and methods. Public Health Nutrition. 2016;19(16):2872-2886.
39. Lee A, Mhurchu CN, Sacks G, et al. Monitoring the price and affordability of foods and diets globally. Obes Rev. Oct 2013;14 Suppl 1:82-95.
40. Yousefian Hansen A, Leighton A, Fox K, Hartley D. Understanding the rural food environment-perspectives of low-income parents. Rural and Remote Health. 2011;11:1631.
41. Jilcott SB, Laraia BA, Evenson KR, Ammerman AS. Perceptions of the community food environment and related influences on food choice among midlife women residing in rural and urban areas: A qualitative analysis. Women & Health. 2009;49(2-3):164-180.
42. Cummins S, Macintyre S. A systematic study of an urban foodscape: The price and availability of food in greater Glasgow. Urban Studies. 2002;39(11):2115-2130.
43. Inglis V, Ball K, Crawford D. Why do women of low socioeconomic status have poorer dietary behaviours than women of higher socioeconomic status? A qualitative exploration. Appetite. 2005;45(3):334-343.

44. Chevalier JA, Kashyap AK, Rossi PE. Why don't prices rise during periods of peak demand? Evidence from scanner data. The American Economic Review. 2003;93(1):15-37.

45. Widener MJ, Metcalf SS, Bar-Yam Y. Dynamic urban food environments a temporal analysis of access to healthy foods. Am J Prev Med. Oct 2011;41(4):439-441.

6.5 Supplemental Tables and Figures

Table 6.8: Average seasonal amplitudes for all products

Product	Amplitude	95% CI	
blueberry	48.12	44.03	52.63
raspberry	36.09	33.27	39.82
blackberry	25.83	22.57	28.69
apricot	14.28	11.93	16.73
strawberry	12.35	11.05	13.92
pomegranate	9.52	7.11	12.34
cherry	8.41	5.72	11.24
asparagus	6.98	5.73	7.98
cucumber	5.73	4.89	6.59
cabbage	5.6	3.46	7.66
pepper	5.53	4.87	6.21
beans	5.45	4.71	6.03
brusselsprouts	5.29	4.31	6.09
squash	4.44	4.18	4.76
grapes	4.29	3.56	5.00
plum	4.24	3.77	4.85
sprouts	4.12	2.89	5.45
peach	4.09	3.38	4.74
peas	4.01	2.91	5.11
nectarine	3.79	2.99	4.63
eggplant	3.5	2.51	4.3
zucchini	3.43	3.01	3.9
vegmix	3.35	2.76	3.94
tomato	2.88	2.21	3.37
parsnip	2.68	1.89	3.4
lettuce	2.65	2.02	3.19
mushroom	2.54	1.76	3.43
cauliflower	2.53	2.17	2.81
endive	2.52	1.81	3.42
radish	2.23	1.54	3.18
bokchoy	2.14	1.58	2.71
kiwi	2.13	1.61	2.64
leek	2.11	1.66	2.7
celery	2.04	1.66	2.37
onion	1.78	1.39	2.15
broccoli	1.6	1.06	2.23
turnip	1.56	1.38	1.79
carrot	1.31	1.08	1.55
potato	1.14	1.02	1.25
banana	1.12	0.61	1.72
watermelon	1.11	0.86	1.4
orange	1.07	0.92	1.25
apple	1.03	0.81	1.25
sweetpotato	1.03	0.82	1.28
beet	0.94	0.68	1.22
papaya	0.91	0.52	1.32
avocado	0.9	0.61	1.14
melon	0.89	0.66	1.12
pineapple	0.86	0.58	1.23
grapefruit	0.79	0.55	1.15
mango	0.78	0.62	0.97

pear	0.69	0.53	0.91

Table 6.9: Average difference in seasonal amplitude between high and low price regions

Product	Difference
pomegranate	5.21
cabbage	4.15
raspberry	4.04
cherry	3.36
banana	1.65
brusselsprouts	1.37
apricot	1.24
beans	1.19
eggplant	1.02
plum	1.00
peas	1.00
leek	0.85
vegmix	0.83
zucchini	0.69
kiwi	0.67
grapefruit	0.64
blueberry	0.62
radish	0.62
sprouts	0.61
lettuce	0.55
pepper	0.54
tomato	0.49
strawberry	0.48
turnip	0.45
papaya	0.31
bokchoy	0.31
orange	0.28
grapes	0.25
carrot	0.25
beet	0.25
parsnip	0.25
endive	0.24
blackberry	0.21
avocado	0.2
broccoli	0.19
celery	0.19
mushroom	0.16
pineapple	0.14
apple	0.14
onion	0.14
nectarine	0.12
melon	0.09
sweetpotato	0.09
squash	0.09
cucumber	0.07
potato	0.07
mango	0.05
watermelon	0.05
peach	0.04
asparagus	0.04
pear	0.01
cauliflower	0.01

Chapter 7 Summary and Conclusions

7.1 Summary of main findings

First, I found that regional and in-store characteristics help to explain the association between accessibility and affordability. While these dimensions were positively correlated in rural areas, the pattern in urban areas was different. Further, the regional chain makeup was an important driver of the association. In my second paper, I found that disparities in the variety and availability of healthy foods aligned with the affordability levels of chains. Lower price chains had a lower proportion of store-months where most of the standard fruit and vegetable basket was available. They also offered lower variety among products that were available. A dose-response relationship between chain affordability and these indicators was evident, but the relationships did not differ based on regional median household income. Third, I found significant seasonal differences in fruit and vegetable prices, and the magnitude of the fluctuations was greater in less affordable regions. Differences in regional and seasonal affordability particularly impact lower income households. Masking this seasonal variation by controlling for season or using only one time point may obscure an important source of disparities.

7.2 Discussion

We may not observe an effect when supermarkets are added to low-income areas because of the under-studied food environment dimensions that are often ignored. Others have observed that treating supermarkets as a homogenous category masks within-store dimensions,[117,123,157] and commented on the potential importance of time.[133] However, affordability, availability and variety have rarely been measured in a large sample of stores or dynamically. I found that the chain make-up and the degree of rurality are important drivers of the association between accessibility and affordability, that availability and variety are lower in low price chains, and substantial seasonal and regional variation in affordability. To the extent that these characteristics matter to consumers, treating supermarkets and time periods as equal masks extensive variation that could render efforts to improve store access fruitless. The heterogeneity could explain, at least in part, previous studies that have found no association between increasing the number of supermarkets and consumption. Further, they help to explain why assuming improved accessibility coincides with improving other food environment dimensions is questionable.[62,70,140,141]

Though supermarkets are not interchangeable, situations without *any* supermarket access may need to be distinguished from those of low access. In areas that entirely lack supermarkets, the impact of adding a store is plausibly more important regardless of within-store attributes. The definition of having ‘no’ access often depends on the personal circumstances that determine mobility.[158] Rural consumers can often travel considerable distance to access supermarkets,[125] though this is not always possible.[127] Urban consumers without cars or public transit might also face access issues, even if their small numbers[17] preclude detecting a general effect of adding a supermarket. In some places, ‘no’ access is more absolute. In Canada, there are remote and First Nations communities that likely have virtually no access. In British Colombia, for instance, people in a northern First Nations community drive five hours to shop at the nearest supermarket in Whitehorse.[159] The local shops’ “limited food variety and high prices,” (pg 3) motivate the drive. One could imagine that the impact of adding any supermarket to such an area could improve local affordability, availability and variety, as well as diet.

Area-level disadvantage does not seem to drive the differences in food environments that foster dietary disparities. The association between area income level and other food environment attributes was not of primary importance in any of my studies, which fails to lend support to the deprivation amplification hypothesis. This is in line with other international research, indicating that this hypothesis rarely holds outside the U.S.[104] To the contrary, deprived areas are often found to have equal or better food environments.[106,115,160-162] Despite relative convergence on this issue, area deprivation is still commonly assumed to drive the food environment disparities that influence dietary choices. While area characteristics could still represent a useful proxy for individuals' socio-economic status when finer data are not available, it is increasingly doubtful that regional deprivation could also proxy the variation in food environments that is thought to contribute to dietary disparities. A stronger theoretical foundation to understand disparities is required.[147,161]

Rather than area-level disadvantage driving food environment disparities, people's differing ability to access what is in their food environment may be more pertinent to understand consumption. Assuming the presence of at least one supermarket, chain make-up is an important driver of regional affordability (Chapter 4). A lack of affordable chains can create one access barrier for low income households. However, where households can access chains within their budgets, we reach a second access barrier. Shopping at a lower-price chain means lower availability and variety of fruits and vegetables (Chapter 5). Opportunities to meet preferences or to expand one's palate are thus fewer. A third barrier is generated through seasonal variation in fruit and vegetable prices (Chapter 6) by reducing consistency in access. Barriers imposed by individuals' interactions with these structural, store-level and temporal factors may also be worsened by social tenets of eating healthy.

Food environments operate within a food culture that could enhance the disparities that these environments drive. This culture has created the perception that healthy eating can be out of reach, regardless of the food environment, even if untrue. The idea perhaps originated in the 1980s when health claims became a marketing tactic. A seminal example is Kellogg's teaming up with the National Cancer Institute to market their wheat bran cereal as protecting against certain forms of cancer. Kellogg's market share increased while the Food and Drug Administration protested the campaign but did not have the power to stop it.[29] Now known as functional foods, demand rose for items marketed using terms like 'low calorie', 'low sodium', and 'low fat.' Functional foods offer the promise to enhance health instead of merely providing the nutrition required to avoid disease. They are symbolically connected with control over one's life, health and well-being.[163] However, not everyone may be able to access these.

Superfoods are a modern example of a functional food that is likely not accessible across socio-economic classes. In the late 2000s, the American Kale Association[•] hired a public relations firm to market kale as the new superfood. On April 22, 2011, Gwyneth Paltrow made kale chips on Ellen and by 2012 kale was Time Magazine's top food trend.[164] Online searches for healthiest vegetables now bring up kale. Avocados have had a similar trajectory. Canadian imports tripled from 25 to 77 thousand metric tons between 2008 and 2016.[165] By 2017, demand overran supply.[166] Calling them one of several 'food trend movements,' Sikka[167] argues that corporations market superfoods as a symbol of opposition to the mass-produced, environmentally and personally harmful products that modern food has come to entail. Other fruits and vegetables that have risen to fame as superfoods include pomegranates, beetroot, blueberries and sweet potatoes in addition to kale. Common to these is that lower price chains are less likely to carry them, and, for example, buying fresh blueberries during the winter is a highly expensive venture. They are thus less accessible to lower income households.

Relying on familiar and cheaper choices does not make eating healthy innately unachievable. However, functional food trends like superfoods make it seem like adherence is necessary for health when it signifies control over one's self and well-being. As Metzl *et al*[168] argue, "health is a term replete with value judgements, hierarchies and blind assumptions that speak as much about power and privilege as they do about well-being."(pg 2) Asserting that unhealthy food is bad for your health is not far from the moralistic view that one is a bad person for eating unhealthy food. As the definition of 'healthy' becomes more classist, so does the opportunity to be 'good.' The emergence of

---

[•] Interestingly, the American Kale Association was found later to not exist. The association was a marketing ploy in itself.

a relatively new eating disorder, *orthorexia nervosa*, may be the culmination. Signs include increasing strictness about the pureness of food and judgement of people who do not eat healthy[169] (pg 48-9). Any divergence from the healthy regime is considered a "fall from grace;"[170,171] an "orthorectic will be plunged into gloom by eating a hot dog, even if his team has just won the world series."[172]

Conflating food culture, its moral connotations and the true meaning of healthy food can result in feelings of failure to achieve what comes to be perceived as an unattainable standard. Though superfoods are not essential, mixed messages from social circles, governments and media create confusion about what is actually healthy.[12] Usual practices suddenly seem 'wrong' and unfamiliar practices are threatening.[12] Consumers may be pushed away from adopting healthier diets altogether because healthy eating becomes just another part of a set of lifestyle activities that are reserved for the elite. For instance, after a calorie-labelling policy passed in New York, an interviewee in a low income neighbourhood explained[173] (pg 456):

> Most people [here] don't care about calories (or) weight loss. They just want to eat, they want the food. The people who care about working out, who go to the gym, who go running marathons, who care about their health…Just those people who be running and jogging, who are thinking about the future and growing up with kids and how they wanna be healthy for their kids and watch their kids grow up – those are the people who care about their health, calories. Most people won't stop and look at the calories.

Individuals actively take up positions contrary to those being advertised[174] because healthy eating is associated with an unavailable level of control over one's destiny.[175] In the extreme, people, "eat what they want to eat, because they just don't care anymore."[173] (pg 456)

Defining true access is complex,[61] as are the beliefs and behaviours that drive consumers' choices.[57] Usually, food environments are viewed as fixed within given areas. Access disparities are then explored as they relate to the food environments in areas with different characteristics. However, disparities may be more easily explored if we consider that individual households are constrained by different budgets and beliefs. Both affect the specific stores they access and the items they are likely to buy. As such, the variation in their experienced food environments rather than area-level disadvantage may be more pertinent to understand access. Different values for food environment dimensions could theoretically be assigned to different households. The dimensions could be thought of as manifestations of disparity rather than as disparities themselves.

Thinking of disparities as occurring at the individual and household level rather than the food environment level underpins the view taken in this thesis. Household economic variation manifests as shopping at different chain price levels, for instance. Accordingly, it is informative to examine food environments in stores with different price levels, as was done in Chapter 5. Economic variation could also manifest as seasonally excluding certain products from individuals' diets. As such, looking at seasonal price differences is informative, as was done in Chapter 6. These ideas hinge on the belief and interest in specific individual-level variation driving food environment disparities. Previous research has already started incorporating affordability into studies of accessibility based on what households of different income levels are likely to access,[123,140,154] however the theory underlying why this should be the case is rarely made explicit.

7.3 Research and Policy Implications

My findings fit within the emerging literature that has questioned the long-held notion that establishing supermarkets in food deserts will improve dietary habits and subsequently health. Despite the local, national and international policy interest, initiatives that target geographic access without considering regional or in-store characteristics are not likely to be effective in improving population diet or in reducing dietary inequalities. I demonstrated that affordability, availability, and variety vary extensively between supermarkets and seasons. Like accessibility, these constructs cannot be considered in isolation. Affordable options are not useful if they do not meet taste preferences. Improved availability and variety are not useful if they are not affordable. And, finally, none of these help if they are not accessible.

To complement existing awareness campaigns, an additional seasonality dimension and controls on marketing could support more fruit and vegetable purchasing across socioeconomic levels. Campaigns have consistently promoted eating a certain number of servings of fruits and vegetables per day.[11] Advertising seasonally-appropriate fruits and vegetables could help consumers to identify what is currently more affordable. To avoid further excluding people, campaigns should advocate nutrition behavior change within existing practices rather than in parallel to them.[12] Promoting simpler, familiar examples of a variety of fruits and vegetables rather than the more extensive variety that may only be available at high price chains or to certain groups could help people to more fully exploit their existing food environments. For example, the Advertising Standards Authority banned an advertisement of superfoods as having clear health benefit.[176]

Further progress in food environments research requires a stronger theoretical foundation and innovative data sources and methods. The focus on aligning food environments with area-level disadvantage continues to dominate the literature. My findings and others' suggest the need for more theory on how food environments foster dietary disparities. Such a foundation requires moving past the deprivation amplification hypothesis as the primary theoretical framework. Innovative data sources and methods could help with this. Previous reviews have often suggested that we need more valid in-store measures.[117] However, new in-store measures will still limit future research possibilities. Using scanner data to measure food environment dimensions is one innovative approach as used in this thesis and in Handbury,[177] though geo-coded Twitter data[178] has also been used to show tweets from so-called food deserts were less likely to have positive sentiments about healthy food. Projects that include visiting large numbers of stores[179] provide an unmatched detailed description of a food environment but they are limited in terms of the possibility to implement over time or space. New data sources have tremendous potential for measuring and monitoring food environments in ways that would be otherwise effectively impossible.

7.4 Strengths and Limitations

This thesis has several strengths. Using marketing data to describe the food environment is innovative and paves the way for similar approaches in the future. To my knowledge, in Canada scanner data has not been used for this purpose, and rarely elsewhere as well. Food environment assessments require such creative means because the necessity to visit stores is prohibitive. Using transactions data enabled fundamental research contributions to this field, for example, availability and variety in particular had never been described on a large spatial scale. Stratification by chains with different price levels and urban and rural environments has also been limited in previous literature. The methods, approach and concepts employed are transferrable to different places and times where such data are available.

In addition, the analytic choices I made in this thesis help to overcome previous methodologic limitations in this literature. A 2015 systematic review on the statistical methods used in food environments literature found that only 14/43 studies accounted for non-normal errors distribution, while 8/43 accounted for the complex nature of clustered data.[180] While some recent developments hold promise,[134,135,181] most studies in this field have not accounted for the potentially incorrect standard error estimates that can result. Using random effects, clustered standard errors and newer methods such as hierarchical bootstrapping,[182] provided a better estimation of precision, and thus reduced the risk of concluding that there are significant associations when there are in fact none.

Despite the strengths, the thesis also has limitations. I focused exclusively on the food environment as it pertains to supermarkets and in high income countries. 'Food swamps' are areas with high numbers of unhealthy outlets, including fast food and convenience stores. They are more prevalent in lower income areas,[16,74,107,142] even in Canada,[162] in contrast to the 'food desert' measure of supermarket density. Unhealthy outlets are also more convincingly associated with adverse consumption and health outcomes.[157] However the topic remains controversial.[183] My findings and discussion also relate mainly to places where supermarkets are broadly considered the best option for a large variety of affordable fruits and vegetables. Supermarkets have different meanings in many places. Elsewhere they are considered sources of processed and junk food, relative to the healthier options available at alternative shopping venues like street markets.[184]

The automatically-collected digital transaction and business directory data used throughout my thesis had weaknesses as well. First, the business directory data were prone to misclassification of store types. I undertook a validation study to mitigate the effect of this limitation (Section 4.3), however even this validation could not have captured stores that were entirely missed. The number of missing stores is not expected to be systematic across different types of areas, however, reducing the potential that our estimates would be biased. Second, the supermarket data were only available at the FSA level creating the potential for misclassification when mapping from FSA centroids to RLS in Paper 1. Exact locations were available for a subset, however, and I found that 90% of the stores' locations mapped to the same RLS as its FSA centroid. Further, the data collection information provided by the marketing company that conducted the sample did not have sufficient detail to employ population weights or make inferences about specific regions. Third, given that we used sales data, products had to be purchased to create a record. There may be bias associated with excluded products, i.e. those that were on offer but not purchased in a given week (Chapters 4 & 6) or month (Chapter 5). Generally, I would expect these items to be more expensive and likely more exclusive.

Finally, this work is ecological and prone to the associated limitations. Minimal variation in prices, supermarket density and mFREI over the years of study made me unable to exploit variation over time for longitudinal analyses, though we still had substantial variation between stores, chains, regions and within years. Reverse causality is unlikely because, within a region, variation in prices, availability and variety occurs exclusively as a result of variation in stores. The opposite cannot be true. Residual confounding is possible. We could not adjust for certain fixed regional and store characteristics such as store size. However, beyond the issues of store misclassification and missing stores as it pertains to the business directory data, there are a limited number of other unobserved factors that occur differentially at a store- or region- level that could affect price, availability or variety. Predicting store attributes is much simpler than trying to explain human behavior.

7.5 Conclusion

Though financing healthy food retailers to locate in underserved areas is a popular policy, food environments research is still at its beginning stages. These retailers, mainly supermarkets, have been treated as a homogenous category. The specific characteristics of these retailers appear to drive the disparities that policy hopes to address, yet these characteristics have rarely been explored. Substantial between-supermarket and seasonal variation in affordability, availability and variety of fruits and vegetables could render efforts to bring just any retailer to any area at any time ineffective. Future directions should include the use of innovative data sources to fully characterize food environment dimensions and developing a stronger theoretical understanding of disparities. Investigating the under-studied food environment dimensions will help to inform policies that could more effectively target dietary disparities and the associated chronic illness and disease.

References

1. U.S. Treasury Department Office of Public Affairs. Obama Administration Details Healthy Food Financing Initiative. 2010; https://www.cdc.gov/chronicdisease/recovery/PDF/Healthy_Food_Financing_release.pdf. Accessed April 1, 2018.

2. Food Access. NYC Food Policy. Accessed April 1, 2018.

3. Philadelphia Department of Public Health. Get Healthy Philly. http://www.phila.gov/health/chronicdisease/gethealthyphilly.html. Accessed April 1, 2018.

4. Hunger Solutions. Good Food Access Fund. http://www.hungersolutions.org/wp-content/uploads/2015/10/GFAF-New-fact-sheet_final02_19_16.pdf. Accessed April 1, 2018.

5. Social Exclusion Unit. Improving shopping access for people living in deprived neighbourhoods. In: Department of Health PAT, ed. Vol 13. London1999.

6. Ahalya M, Jane Y P, Éric R, Marc L, Tina M, Leia M M. Geographic retail food environment measures for use in public health. *Health Promotion and Chronic Disease Prevention in Canada : Research, Policy and Practice*. 2017;37(10):357-362.

7. City of Vancouver. *Vancouver Food Strategy*. 2013.

8. Toronto Public Health. *Toronto Food Strategy: 2015 Update*. 2015.

9. Micha R, Khatibzadeh S, Shi P, Andrews KG, Engell RE, Mozaffarian D. Global, regional and national consumption of major food groups in 1990 and 2010: a systematic analysis including 266 country-specific nutrition surveys worldwide. *BMJ Open*. 2015;5(9).

10. Lock K, Pomerleau J, Causer L, Altmann DR, McKee M. The global burden of disease attributable to low consumption of fruit and vegetables: implications for the global strategy on diet. *Bulletin of the World Health Organization*. 2005;83(2):100-108.

11. Rekhy R, McConchie R. Promoting consumption of fruit and vegetables for better health. Have campaigns delivered on the goals? *Appetite*. 2014;79:113-123.

12. Steven MC. Strategies to influence nutrition behaviour. *Nutrition and Health*. 1990;6(4):189-204.

13. Dixon H, Borland R, Segan C, Stafford H, Sindall C. Public reaction to Victoria's “2 Fruit ‘n’ 5 Veg Every Day” campaign and reported consumption of fruit and vegetables. *Preventive Medicine*. 1998;27(4):572-582.

14. Mozaffarian D, Afshin A, Benowitz NL, et al. Population approaches to improve diet, physical activity, and smoking Habits: A scientific statement from the American Heart Association. *Circulation*. 2012;CIR-0b013e318260a20b.

15. Glanz K, Sallis JF, Saelens BE, Frank LD. Healthy nutrition environments: concepts and measures. *Am J Health Promot*. May-Jun 2005;19(5):330-333.

16. Lee H. The role of local food availability in explaining obesity risk among young school-aged children. *Social Science & Medicine*. 2012;74(8):1193-1203.

17. Block JP, Subramanian SV. Moving beyond "Food Deserts": Reorienting United States policies to reduce disparities in diet quality. *Plos Medicine*. Dec 2015;12(12).

18. Larsen K, Gilliland J. A farmers’ market in a food desert: Evaluating impacts on the price and availability of healthy food. *Health & Place*. 2009;15(4):1158-1162.

19. Andreyeva T, Blumenthal DM, Schwartz MB, Long MW, Brownell KD. Availability and prices of foods across stores and neighborhoods: The case of New Haven, Connecticut. *Health Aff (Millwood)*. Sep-Oct 2008;27(5):1381-1388.

20. Wheelock V. *Implementing Dietary Guidelines for Healthy Eating*: Blackie Academic & Professional; 1997.

21. Willett W. Diet and health: What should we eat? *Science (New York, N.Y.)*. 1994;264(5158):532-537.

22. Strong K, Mathers C, Leeder S, Beaglehole R. Preventing chronic diseases: how many lives can we save? *The Lancet*. 2005;366(9496):1578-1582.

23. Gonzalez CA, Riboli E. Diet and cancer prevention: Contributions from the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *European Journal of Cancer*. 2010;46(14):2555-2562.

24. Jacka FN, Kremer PJ, Berk M, et al. A prospective study of diet quality and mental health in adolescents. *PloS One*. 2011;6(9):e24805.

25. Freeman LR, Haley-Zitlin V, Rosenberger DS, Granholm A-C. Damaging effects of a high-fat diet to the brain and cognition: A review of proposed mechanisms. *Nutritional Neuroscience*. 2014;17(6):241-251.

26. Lustig RH, Schmidt LA, Brindis CD. The toxic truth about sugar. *Nature*. 2012;482:27.

27. Ouyang X, Cirillo P, Sautin Y, et al. Fructose consumption as a risk factor for non-alcoholic fatty liver disease. *Journal of Hepatology*. 2008;48(6):993-999.

28. Beilharz J, Maniam J, Morris M. Diet-induced cognitive deficits: The role of fat and sugar, potential mechanisms and nutritional interventions. *Nutrients*. 2015;7(8):5307.

29. Sims LS. Government involvement in nutrition education: panacea or Pandora's box? *Health Education Research*. 1990;5(4):517-526.

30. Anderson JW, Baird P, Davis JRH, et al. Health benefits of dietary fiber. *Nutrition Reviews*. 2009;67(4):188-205.

31. Liu RH. Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. *The American Journal of Clinical Nutrition*. 2003;78(3):517S-520S.

32. Prior RL, Cao G. Antioxidant phytochemicals in fruits and vegetables: diet and health implications. *HortScience*. 2000;35(4):588-592.

33. World Health Organization. *Promoting fruit and vegetable consumption around the world*2013.

34. Reiss R, Johnston J, Tucker K, DeSesso JM, Keen CL. Estimation of cancer risks and benefits associated with a potential increased consumption of fruits and vegetables. *Food and Chemical Toxicology*. 2012;50(12):4421-4427.

35. Soerjomataram I, Oomen D, Lemmens V, et al. Increased consumption of fruit and vegetables and future cancer incidence in selected European countries. *European Journal of Cancer*. 2010;46(14):2563-2580.

36. Patterson BH, Block G, Rosenberger WF, Pee D, Kahle LL. Fruit and vegetables in the American diet: data from the NHANES II survey. *American Journal of Public Health*. 1990;80(12):1443-1449.

37. Minister of Industry. *Healthy people, healthy places*. 2010.

38. Colapinto CK, Graham J, St-Pierre S. Trends and correlates of frequency of fruit and vegetable consumption, 2007 to 2014. *Health Reports*. 2018;29(1):9-14.

39. Mooney C. Cost and availability of healthy food choices in a London health district. *Journal of Human Nutrition and Dietetics*. 1990;3(2):111-120.

40. Casagrande SS, Wang Y, Anderson C, Gary TL. Have Americans increased their fruit and vegetable intake? *Am J Prev Med*. 2007;32(4):257-263.

41. Ashfield-Watt PA. Fruits and vegetables, 5+ a day: are we getting the message across? *Asia Pacific Journal of Clinical Nutrition*. 2006;15(2):245-252.

42. Paquette M-C. Perceptions of healthy eating: state of knowledge and research gaps. *Canadian Journal of Public Health/Revue Canadienne de Santé Publique*. 2005:S15-S19.

43. 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 and Nutrition Bulletin*. 2011;32(4):365-375.

44. Sniehotta FF, Presseau J, Araújo-Soares V. Time to retire the theory of planned behaviour. *Health Psychology Review*. 2014;8(1):1-7.

45. Ajzen I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*. 1991;50(2):179-211.

46. Godin G, Kok G. The Theory of Planned Behavior: A review of its applications to health-related behaviors. *American Journal of Health Promotion*. 1996;11(2):87-98.

47. Boyd-Orr J. *Food, Health and Income: Report on a Survey of Adequacy of Diet in Relation to Income*. London: Macmillan & Co; 1937.

48. Smith A, Smith C. Dietary intake and lifestyle patterns: correlates with socio-economic, demographic and environmental factors. *Journal of Human Nutrition and Dietetics*. 1994;7(4):283-294.

49. Roos G, Johansson L, Kasmel A, Klumbiené J, Prättälä R. Disparities in vegetable and fruit consumption: European cases from the north to the south. *Public Health Nutrition*. 2007;4(1):35-43.

50. Grimm KA, Foltz JL, Blanck HM, Scanlon KS. Household Income disparities in fruit and vegetable consumption by state and territory: Results of the 2009 Behavioral Risk Factor Surveillance System. *Journal of the Academy of Nutrition and Dietetics*. 2012;112(12):2014-2021.

51. Dehghan M, Akhtar-Danesh N, Merchant AT. Factors associated with fruit and vegetable consumption among adults. *Journal of Human Nutrition and Dietetics*. 2011;24(2):128-134.

52. Azagba S, Sharaf MF. Disparities in the frequency of fruit and vegetable consumption by socio-demographic and lifestyle characteristics in Canada. *Nutrition Journal*. 2011;10(1):118.

53. van Lenthe FJ, Jansen T, Kamphuis CB. Understanding socio-economic inequalities in food choice behaviour: can Maslow's pyramid help? *British Journal of Nutrition*. 2015;113(07):1139-1147.

54. Sallis JO, Neville. Ecological models of health behaviour. In: Glanz KR, BK; Viswanath, K, ed. *Health Behavior: Theory, Research, and Practice*. Vol 5th Edition. San Francisco, CA: Jossey-Bass; 2015.

55. Cummins S, Macintyre S. Food environments and obesity—neighbourhood or nation? *International Journal of Epidemiology*. 2006;35(1):100-104.

56. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating healthy food and eating environments: policy and environmental approaches. *Annu. Rev. Public Health*. 2008;29:253-272.

57. Glanz K, Mullis RM. Environmental interventions to promote healthy eating: a review of models, programs, and evidence. *Health Education Quarterly*. 1988;15(4):395-415.

58. Cole-Hamilton I, Gunner K, Leverkus C, Starr J. A study among dietitians and adult members of their households of the practicalities and implications of following proposed dietary guidelines for the UK. British Dietetic Association Community Nutrition Group Nutrition Guidelines Project. *Human nutrition. Applied nutrition*. 1986;40(5):365-389.

59. McCarthy EJ. *Basic Marketing: A Managerial Approach*: R.D. Irwin; 1960.

60. Sharkey JD, WE; Nalty, C,. Convenience Stores and the Marketing of Foods and Beverages Through Product Assortment. *Am J Prev Med*. 2012;43(3):S109-S115.

61. Charreire H, Casey R, Salze P, et al. Measuring the food environment using geographical information systems: A methodological review. *Public Health Nutrition*. 2010;13(11):1773-1785.

62. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: A systematic review. *Health & Place*. 2012;18(5):1172-1187.

63. Greer S, Schieb L, Schwartz G, Onufrak S, Park S. Association of the neighborhood retail food environment with sodium and potassium Intake among US Adults. *Preventing Chronic Disease*. 2014;11:E70.

64. Centres for Disease Control and Prevention. Children's Food Environment State Indicator Report. In: Health and Human Services, ed2011.

65. Cummins S, Findlay A, Petticrew M, Sparks L. Retail-led regeneration and store-switching behaviour. *Journal of Retailing and Consumer Services*. 2008;15(4):288-295.

66. Dubowitz T, Ghosh-Dastidar M, Cohen DA, et al. Diet and perceptions change with supermarket introduction in a food desert, but not because of supermarket use. *Health Affairs*. 2015;34(11):1858-1868.

67. 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 Nutrition*. 2015;18(15):2881-2890.

68. Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS. The built environment and obesity: A systematic review of the epidemiologic evidence. *Health & Place*. 2010;16(2):175-190.

69. Holsten JE. Obesity and the community food environment: a systematic review. *Public Health Nutrition*. 2009;12(3):397-405.

70. Moore LV, Diez Roux AV, Nettleton JA, Jacobs Jr DR. Associations of the local food environment with diet quality—a comparison of assessments based on surveys and geographic information systems: the multi-ethnic study of atherosclerosis. *American Journal of Epidemiology*. 2008;167(8):917-924.

71. Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obesity Reviews*. 2011;12(5):e95-e106.

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

73. Lopez RP. Neighborhood risk factors for obesity. *Obesity*. 2007;15(8):2111-2119.

74. Bodor JN, Rice JC, Farley TA, Swalm CM, Rose D. The association between obesity and urban food environments. *Journal of Urban Health*. 2010;87(5):771-781.

75. Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obesity Reviews*. 2011;12(3):217-230.

76. Rose D, Hutchinson PL, Bodor JN, et al. Neighborhood food environments and body mass index:The importance of in-store contents. *Am J Prev Med*. Sep 2009;37(3):214-219.

77. Cassady D, Jetter KM, Culp J. Is price a barrier to eating more fruits and vegetables for low-income families? *Journal of the American Dietetic Association*. 2007;107(11):1909-1915.

78. Monsivais P, Aggarwal A, Drewnowski A. Following federal guidelines to increase nutrient consumption may lead to higher food costs for consumers. *Health Affairs*. 2011;30(8):1471-1477.

79. Horning ML, Fulkerson JA. A systematic review on the affordability of a healthful diet for families in the United States. *Public Health Nursing*. Jan-Feb 2015;32(1):68-80.

80. Rao M, Afshin A, Singh G, Mozaffarian D. Do healthier foods and diet patterns cost more than less healthy options? A systematic review and meta-analysis. *BMJ Open*. 2013;3(12):e004277.

81. Monsivais P, McLain J, Drewnowski A. The rising disparity in the price of healthful foods: 2004–2008. *Food Policy*. 2010;35(6):514-520.

82. Hakansson A. Has it become increasingly expensive to follow a nutritious diet? Insights from a new price index for nutritious diets in Sweden 1980-2012. *Food Nutr Res*. 2015;59:26932.

83. Niebylski ML, Redburn KA, Duhaney T, Campbell NR. Healthy food subsidies and unhealthy food taxation: A systematic review of the evidence. *Nutrition*. 2015;31(6):787-795.

84. Nnoaham KE, Sacks G, Rayner M, Mytton O, Gray A. Modelling income group differences in the health and economic impacts of targeted food taxes and subsidies. *International Journal of Epidemiology*. 2009:dyp214.

85. Andreyeva T, Long MW, Brownell KD. The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food. *American Journal of Public Health*. 2010;100(2):216-222.

86. Sturm R, Datar A. Body mass index in elementary school children, metropolitan area food prices and food outlet density. *Public Health*. 2005;119(12):1059-1068.

87. Anekwe TD, Rahkovsky I. The Association Between Food Prices and the Blood Glucose Level of US Adults With Type 2 Diabetes. *American Journal of Public Health*. 2014;104(4):678-685.

88. Ard JD, Fitzpatrick S, Desmond RA, et al. The impact of cost on the availability of fruits and vegetables in the homes of schoolchildren in Birmingham, Alabama. *American Journal of Public Health*. 2007;97(2):367-372.

89. Gelbach JB, Klick J, Stratmann T. Cheap donuts and expensive broccoli: the effect of relative prices on obesity. *Available at SSRN 976484*. 2009.

90. Stigler GJ. The cost of subsistence. *Journal of Farm Economics*. 1945;27(2):303-314.

91. Darmon N, Ferguson E, Briend A. Do economic constraints encourage the selection of energy dense diets? *Appetite*. 2003;41(3):315-322.

92. Cummins S, Smith DM, Aitken Z, et al. Neighbourhood deprivation and the price and availability of fruit and vegetables in Scotland. *Journal of Human Nutrition and Dietetics*. Oct 2010;23(5):494-501.

93. Darmon N, Ferguson EL, Briend A. Impact of a cost constraint on nutritionally adequate food choices for French women: an analysis by linear programming. *Journal of Nutrition Education and Behavior*. 2006;38(2):82-90.

94. Maillot M, Darmon N, Drewnowski A. Are the lowest-cost healthful food plans culturally and socially acceptable? *Public Health Nutrition*. Aug 2010;13(8):1178-1185.

95. Pollard J, Kirk SFL, Cade JE. Factors affecting food choice in relation to fruit and vegetable intake: a review. *Nutrition Research Reviews*. 2013;15(2):373-387.

96. Black C, Ntani G, Kenny R, et al. Variety and quality of healthy foods differ according to neighbourhood deprivation. *Health & Place*. Nov 2012;18(6):1292-1299.

97. Ton Z, Raman A. The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production and Operations Management*. 2010;19(5):546-560.

98. Iyengar SS, Lepper MR. When choice is demotivating: can one desire too much of a good thing? *Journal of Personality and Social Psychology*. Dec 2000;79(6):995-1006.

99. Curhan RC. The effects of merchandising and temporary promotional activities on the sales of freshfruits and vegetables in supermarkets. *Journal of Marketing Research*. 1974;11(3):286-294.

100. van Kleef E, Otten K, van Trijp HC. Healthy snacks at the checkout counter: a lab and field study on the impact of shelf arrangement and assortment structure on consumer choices. *BMC Public Health*. 2012;12:1072.

101. Martin KS, Havens E, Boyle KE, et al. If you stock it, will they buy it? Healthy food availability and customer purchasing behaviour within corner stores in Hartford, CT, USA. *Public health nutrition*. Oct 2012;15(10):1973-1978.

102. Izumi BT, Zenk SN, Schulz AJ, Mentz GB, Wilson C. Associations between neighborhood availability and individual consumption of dark-green and orange Vegetables among ethnically diverse adults in Detroit. *Journal of the American Dietetic Association*. 2011;111(2):274-279.

103. Franco M, Diez-Roux AV, Nettleton JA, et al. Availability of healthy foods and dietary patterns: the Multi-Ethnic Study of Atherosclerosis. *The American Journal of Clinical Nutrition*. 2009;89(3):897-904.

104. Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? *International Journal of Behavioral Nutrition and Physical Activity*. 2007;4(1):32.

105. Beaulac J, Kristjansson E, Cummins S. A systematic review of food deserts, 1966-2007. *Preventing Chronic Disease*. 2009;6(3):A105.

106. Smith DM, Cummins S, Taylor M, et al. Neighbourhood food environment and area deprivation: spatial accessibility to grocery stores selling fresh fruit and vegetables in urban and rural settings. *International Journal of Epidemiology*. 2010;39(1):277-284.

107. Latham J, Moffat T. Determinants of variation in food cost and availability in two socioeconomically contrasting neighbourhoods of Hamilton, Ontario, Canada. *Health & Place*. Mar 2007;13(1):273-287.

108. Larsen K, Gilliland J. Mapping the evolution of 'food deserts' in a Canadian city: Supermarket accessibility in London, Ontario, 1961–2005. *International Journal of Health Geographics*. 2008;7(1):16.

109. Apparicio P, Cloutier M-S, Shearmur R. The case of Montreal's missing food deserts: Evaluation of accessibility to food supermarkets. *International Journal of Health Geographics*. 2007;6(1):1.

110. Pouliot N, Hamelin A-M. Disparities in fruit and vegetable supply: a potential health concern in the greater Quebec City area. *Public Health Nutrition*. 2009;12(11):2051-2059.

111. Gould AC, Apparicio P, Cloutier MS. Classifying neighbourhoods by level of access to stores selling fresh fruit and vegetables and groceries: Identifying problematic areas in the City of Gatineau, Quebec. *Canadian Journal of Public Health-Revue Canadienne De Santé Publique*. Nov-Dec 2012;103(6):E433-E437.

112. Polsky JY, Moineddin R, Glazier RH, Dunn JR, Booth GL. Foodscapes of southern Ontario: neighbourhood deprivation and access to healthy and unhealthy food retail. *Canadian Journal of Public Health-Revue Canadienne De Santé Publique*. Sep-Oct 2014;105(5):e369-375.

113. Smoyer-Tomic KE, Spence JC, Amrhein C. Food deserts in the prairies? Supermarket accessibility and neighborhood need in Edmonton, Canada. *The Professional Geographer*. 2006;58(3):307-326.

114. Wang J, Engler-Stringer R, Muhajarine N. Assessing the consumer food environment in restaurants by neighbourhood distress level across Saskatoon, Saskatchewan. *Canadian Journal of Dietetic Practice and Research*. 2015;77(1):9-16.

115. Cummins S, Macintyre S. “Food deserts”—evidence and assumption in health policy making. *BMJ*. 2002;325(7361):436-438.

116. Mendoza RU. Why do the poor pay more? Exploring the poverty penalty concept. *Journal of International Development*. 2011;23(1):1-28.

117. Gustafson A, Hankins S, Jilcott S. Measures of the consumer food store environment: A systematic review of the evidence 2000–2011. *Journal of Community Health*. 2012;37(4):897-911.

118. Ball K, Timperio A, Crawford D. Neighbourhood socioeconomic inequalities in food access and affordability. *Health & Place*. 2009;15(2):578-585.

119. Block D, Kouba J. A comparison of the availability and affordability of a market basket in two communities in the Chicago area. *Public Health Nutrition*. 2006;9(07):837-845.

120. de Roos B, Binacchi F, Whybrow S, Sneddon AA. Differences in expenditure and amounts of fresh foods, fruits and vegetables, and fish purchased in urban and rural Scotland. *Public Health Nutrition*. Feb 2017;20(3):524-533.

121. Cheadle A, Psaty BM, Curry S, et al. Community-level comparisons between the grocery store environment and individual dietary practices. *Preventive Medicine*. 1991;20(2):250-261.

122. Sloane DC, Diamant AL, Lewis LB, et al. Improving the nutritional resource environment for healthy living through community-based participatory research. *Journal of General Internal Medicine*. 2003;18(7):568-575.

123. Black C, Ntani G, Inskip H, et al. Measuring the healthfulness of food retail stores: variations by store type and neighbourhood deprivation. *International Journal of Behavioral Nutrition and Physical Activity*. 2014;11(1):69.

124. Hartley D, Anderson N, Fox K, Lenardson J. How does the rural food environment affect rural childhood obesity? *Childhood Obesity*. 2011;7(6):450-461.

125. Smith C, Miller H. Accessing the food systems in urban and rural Minnesotan communities. *J Nutr Educ Behav*. Nov-Dec 2011;43(6):492-504.

126. Yousefian Hansen A, Leighton A, Fox K, Hartley D. Understanding the rural food environment-perspectives of low-income parents. *Rural and Remote Health*. 2011;11:1631.

127. Lucan SC, Gustafson A, Jilcott Pitts SB. The concept of “rural food deserts” is still meaningful. *Childhood Obesity*. 2012;8(5):484-485.

128. Pinard CA, Shanks CB, Harden S, Yaroch A. An integrative literature review of small food store research across urban and rural communities in the US. *Preventive Medicine Reports.* 2016;3:324-332.

129. Hendrickson D, Smith C, Eikenberry N. Fruit and vegetable access in four low-income food deserts communities in Minnesota. *Agriculture and Human Values.* 2006;23(3):371-383.

130. Byker Shanks C, Ahmed S, Smith T, et al. Availability, price, and quality of fruits and vegetables in 12 rural Montana counties. *Preventing Chronic Disease.* 2015;12:E128.

131. Hardin-Fanning F, Rayens MK. Food cost disparities in rural communities. *Health Promot Pract.* May 2015;16(3):383-391.

132. Lenardson JD, Hansen AY, Hartley D. Rural and remote food environments and obesity. *Current Obesity Reports.* 2015;4(1):46-53.

133. Widener MJ, Shannon J. When are food deserts? Integrating time into research on food accessibility. *Health & Place.* Nov 2014;30:1-3.

134. Lamichhane AP, Warren JL, Peterson M, Rummo P, Gordon-Larsen P. Spatial-temporal modeling of neighborhood sociodemographic characteristics and food stores. *American Journal of Epidemiology.* Jan 2015;181(2):137-150.

135. Luan H, Law J, Quick M. Identifying food deserts and swamps based on relative healthy food access: a spatio-temporal Bayesian approach. *International Journal of Health Geographics.* Dec 2015;14.

136. Filomena S, Scanlin K, Morland KB. Brooklyn, New York foodscape 2007–2011: a five-year analysis of stability in food retail environments. *International Journal of Behavioral Nutrition and Physical Activity.* 2013;10(1):46.

137. Widener MJ, Metcalf SS, Bar-Yam Y. Dynamic urban food environments a temporal analysis of access to healthy foods. *Am J Prev Med.* Oct 2011;41(4):439-441.

138. Valpiani N, Wilde P, Rogers B, Stewart H. Patterns of fruit and vegetable availability and price competitiveness across four seasons are different in local food outlets and supermarkets. *Public Health Nutrition.* Oct 2015;18(15):2846-2854.

139. Ma Y, Labban A, Cherian M, Shaban-Nejad A, Buckeridge DL, Dubé L. System of indicators for the nutritional quality of marketing and food environment: product quality, availability, affordability, and promotion. *Diet Quality*: Springer; 2013:383-396.

140. Jiao J, Moudon AV, Ulmer J, Hurvitz PM, Drewnowski A. How to identify food deserts: Measuring physical and economic access to supermarkets in King County, Washington. *American Journal of Public Health.* 2012;102(10):e32-e39.

141. Kelly B, Flood VM, Yeatman H. Measuring local food environments: an overview of available methods and measures. *Health & Place.* Nov 2011;17(6):1284-1293.

142. Cummins S, Macintyre S. A systematic study of an urban foodscape: The price and availability of food in Greater Glasgow. *Urban Studies.* 2002;39(11):2115-2130.

143. Black C, Moon G, Baird J. Dietary inequalities: What is the evidence for the effect of the neighbourhood food environment? *Health & Place.* 2014;27:229-242.

144. Black JL, Carpiano RM, Fleming S, Lauster N. Exploring the distribution of food stores in British Columbia: Associations with neighbourhood socio-demographic factors and urban form. *Health & Place.* 2011;17(4):961-970.

145. Kincses Á, Tóth G. The application of gravity model in the investigation of spatial structure. *Acta Polytechnica Hungarica.* 2014;11(2):5-19.

146. Zipf GK. *Human behavior and the principle of least effort: an introduction to human ecology*: Addison-Wesley Press; 1949.

147. Cummins S. Commentary: Investigating neighbourhood effects on health—avoiding the 'Local Trap'. *International Journal of Epidemiology.* 2007;36(2):355-357.

148. Kestens Y, Lebel A, Daniel M, Thériault M, Pampalon R. Using experienced activity spaces to measure foodscape exposure. *Health & Place.* 2010;16(6):1094-1103.

149. Gustafson A, Christian JW, Lewis S, Moore K, Jilcott S. Food venue choice, consumer food environment, but not food venue availability within daily travel patterns are associated with dietary intake among adults, Lexington Kentucky 2011. *Nutrition Journal.* 2013;12(1):17.

150. Dubowitz T, Zenk SN, Ghosh-Dastidar B, et al. Healthy food access for urban food desert residents: examination of the food environment, food purchasing practices, diet and BMI. *Public Health Nutrition.* Aug 2015;18(12):2220-2230.

151. Drewnowski A, Aggarwal A, Hurvitz PM, Monsivais P, Moudon AV. Obesity and supermarket access: proximity or price? *American Journal of Public Health.* 2012;102(8):e74-e80.

152. Glanz K, Sallis JF, Saelens BE, Frank LD. Nutrition Environment Measures Survey in Stores (NEMS-S). *Am J Prev Med.* 2007;32(4):282-289.

153. Ni Mhurchu C, Eyles H, Schilling C, et al. Food Prices and consumer demand: Differences across income Levels and ethnic groups. *PLoS One.* 2013;8(10):e75934.

154. Mackenbach JD, Burgoine T, Lakerveld J, et al. Accessibility and affordability of supermarkets: Associations with the DASH diet. *Am J Prev Med.* 2017;53(1):55-62.

155. Ghosh-Dastidar B, Cohen D, Hunter G, et al. Distance to store, food prices, and obesity in urban food deserts. *Am J Prev Med.* Nov 2014;47(5):587-595.

156. Thornton LE, Kavanagh AM. Understanding the local food environment and obesity. In: Witten K, ed. *Geographies of Obesity Environmental Understandings of the Obesity Epidemic*. New York: Routledge; 2010.

157. 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. *Archives of Internal Medicine.* 2011;171(13):1162-1170.

158. Williams P, Hubbard P. Who is disadvantaged? Retail change and social exclusion. *The International Review of Retail, Distribution and Consumer Research.* 2001;11(3):267-286.

159. Dietitians of Canada. The cost of eating in BC 2009: Low income British Columbians can't afford healthy food. *Community Nutritionists Council of BC.* 2009.

160. Mercille G, Richard L, Gauvin L, Kestens Y, Payette H, Daniel M. Comparison of two Indices of availability of fruits/vegetable and fast food outlets. *Journal of Urban Health.* 2013;90(2):240-245.

161. Macintyre S, Macdonald L, Ellaway A. Do poorer people have poorer access to local resources and facilities? The distribution of local resources by area deprivation in Glasgow, Scotland. *Social Science & Medicine.* 2008;67(6):900-914.

162. Health Canada. Measuring the Food Environment in Canada. *www.hc-sc.gc.ca/fn-an/nutrition/pol/index-eng.php.* 2013.

163. Urala N, Lähteenmäki L. Reasons behind consumers' functional food choices. *Nutrition & Food Science.* 2003;33(4):148-158.

164. Gustafson E. How Kale Became Cool. *SELF.* March 27, 2015. https://www.self.com/story/how-kale-became-cool.

165. StatCan. Volume of fresh avocados imported in Canada from 2008 to 2016 (in thousand metric tons). *Statista - The Statistics Portal.* (n.d.); Retrieved May 21, 2018, from https://www.statista.com/statistics/453463/fresh-avocado-import-volume-canada/.

166. Parkin B. Holy Guacamole! Avocado Prices Rise to Record Highs. *The Wall Street Journal.* Sep 6, 2017.

167. Sikka T. Contemporary superfood cults: nutritionism, neoliberalism, and gender. In: Cargill K, ed. *Food Cults: How Fads, Dogma, and Doctrine Influence Diet*. London: Rowman & Littlefield; 2017:87-108.

168. Metzl JM, Kirkland A. *Against Health: How Health Became the New Morality*: NYU Press; 2010.

169. Bratman SK, D. *Health Food Junkies: The Rise of Orthorexia Nervosa - the Health Food Eating Disorder*: Broadway Books; 2004.

170. Varga M, Dukay-Szabo S, Tury F, van Furth EF. Evidence and gaps in the literature on orthorexia nervosa. *Eating and Weight Disorders : EWD.* Jun 2013;18(2):103-111.

171. Donini LM, Marsili D, Graziani MP, Imbriale M, Cannella C. Orthorexia nervosa: a preliminary study with a proposal for diagnosis and an attempt to measure the dimension of the phenomenon. *Eating and Weight Disorders : EWD*. Jun 2004;9(2):151-157.

172. Bratman S. The Health Food Eating Disorder. *Yoga*: Taken from: http://www.orthorexia.com/original-orthorexia-essay/ 1997.

173. Gordon C, Hayes R. Counting calories: Resident perspectives on calorie labeling in New York City. *Journal of Nutrition Education and Behavior*. 2012;44(5):454-458.

174. Rhodewalt F, Davison J. Reactance and the coronary-prone behavior pattern: the role of self-attribution in responses to reduced behavioral freedom. *Journal of Personality and Social Psychology*. 1983;44(1):220.

175. Gough B, Conner MT. Barriers to healthy eating amongst men: a qualitative analysis. *Social Science & Medicine*. 2006;62(2):387-395.

176. Advertising Standards Authority. ASA Adjudication on Pharmacare (Europe) Ltd t/a Bioglan. *https://www.asa.org.uk/rulings/pharmacare-europe-ltd-a14-260406.html#.VBBdnGRdWrG* 2014. Accessed May 18, 2018.

177. Handbury J, Rahkovsky I, Schnell M. *Is the focus on food deserts fruitless? Retail access and food purchases across the socioeconomic spectrum*: National Bureau of Economic Research;2015.

178. Widener MJ, Li W. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography*. 2014;54:189-197.

179. Minaker LM, Olstad DL, Thompson ME, Raine KD, Fisher P, Frank LD. Associations between frequency of food shopping at different store types and diet and weight outcomes: findings from the NEWPATH study. *Public Health Nutrition*. 2016;19(12):2268-2277.

180. Lamb KE, Thornton LE, Cerin E, Ball K. Statistical approaches used to assess the equity of access to food outlets: A systematic review. *AIMS Public Health*. 2015;2(3):358-401.

181. Luan H, Minaker LM, Law J. Do marginalized neighbourhoods have less healthy retail food environments? An analysis using Bayesian spatial latent factor and hurdle models. *International Journal of Health Geographics*. 2016;15(1):29.

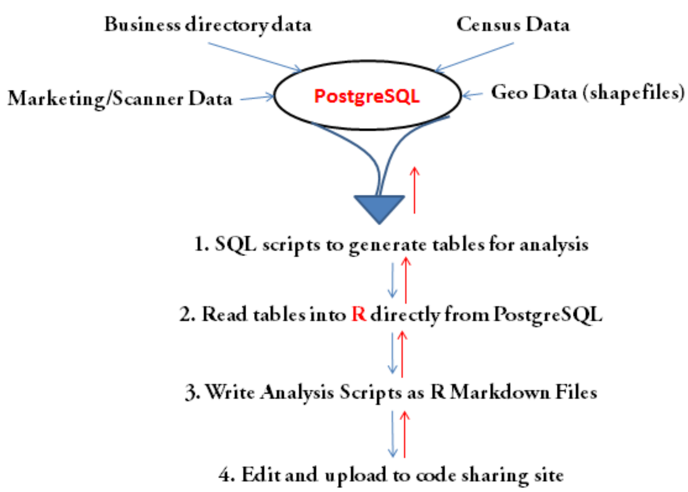
182. Ren S, Lai H, Tong W, Aminzadeh M, Hou X, Lai S. Nonparametric bootstrapping for hierarchical data. *Journal of Applied Statistics*. 2010;37(9).

183. Sturm R, Hattori A. Diet and obesity in Los Angeles County 2007–2012: Is there a measurable effect of the 2008 "Fast-Food Ban"? *Social Science & Medicine*. 2015;133:205-211.

184. Asfaw A. Does supermarket purchase affect the dietary practices of households? Some empirical evidence from Guatemala. *Development Policy Review*. 2008;26(2):227-243.

Supplementary Information: Data Management & Reproducibility

Working with scanner data alongside several other data sources required particular mitigation techniques to ensure data management was appropriate and reproducibility was prioritized. Data management was a central challenge in this thesis. Due to its size, it was not possible to store the data in the same software used to analyse it, or to handle all the data at once. The raw data were thus quite disjointed, required various levels of processing at different stages, as well as simple ways to identify reasons for outliers or other data anomalies. To manage this, I created a direct workflow from the raw data to analysis and vice versa:



1. PostgreSQL was used to store all the data and do higher-level data structuring, including generating the price variables and geocoding the stores in the data.
2. Once the tables were more manageable, I read their relevant components into R. Here, I undertook further processing that required more complex operations such as aggregation.
3. Analysis relied on the data generated from the R script in Step 2. No further general data processing was done at this stage to avoid errors.
4. The analysis script was edited for clarity and shared on the popular version control site, Github.

In reverse, issues found during analysis could be explored by following the workflow from the analysis to the SQL script to the raw data. In this way, the data remained organized and each analysis was clearly linked to the original data. The final coding files are available here: https://github.com/deepajag/food-prices.