Characterizing Community Dietary Patterns Using Grocery Point-of-Sales Data

Hiroshi Mamiya

Department of Epidemiology, Biostatistics, and Occupational Health

McGill University, Montreal, Canada

March 2020

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

Abstract

Background

Unhealthy diet is the most important preventable risk factor globally for premature death and disability. Regional nutrition inequality is expected to exist across towns or even within city subdivisions due to geographic variation in socio-cultural, economic, marketing and built environment features. Effective development and evaluation of nutrition policies and community interventions rely on the availability of measures capturing geographic patterns of diets and environmental factors influencing food selection. A traditional data source for nutrition surveillance is population health surveys, which have inadequate sample size to provide stable estimates of community diets. Furthermore, surveys are unable to measure important within-store marketing of unhealthy food, such as price discounting, which may have important, local-in-time effects. Electronic Point-Of-Sales (POS, i.e. store- rather than individual shopper-level) grocery transaction records generated by product scanning have the potential to overcome these limitations with appropriate analytical approaches.

Objectives

The overarching goal of this thesis is to develop and evaluate measurement methods that improve nutrition surveillance, specifically focusing on public health indicators of within-store food marketing (Objective1) and community diets with a high spatial resolution (Objectives 2 and 3):

Objective 1: To estimate how neighbourhood education modifies the effect of price discounting on soda purchasing at the store level.

Objective 2**:** To develop a small-area indicator of latent soda demand using a hierarchical Bayesian model and to evaluate the ability of the indicator to explain area-level risk of type 2 diabetes mellitus.

Objective 3**:** To develop small-area indicators of soda, diet-soda, sugary yogurt and plain yogurt using a retail gravity model and to evaluate the ability of these indicators to explain area-level risk of type 2 diabetes mellitus.

Methods and Results

Objective 1: I modelled weekly store-level sales of soda as a function of store-level price discounting, store- and area-level confounders, and an interaction term between discounting and categorical area-level education in the area where stores were located. The estimated effect of discounting on soda sales was larger in stores in less educated neighbourhoods, most notably in pharmacies.

Objective 2: I modelled the store-level proportion of soda sales out of the total sales of all beverages as a function of store- and area-level covariates using a hierarchical Bayesian model with a random effect specified by a proper Conditional Autoregressive (CAR) prior. The estimated posterior mean of the CAR prior is the proposed area-level indicator of latent soda demand and is estimable for all areas through smoothing, including in areas without stores. To evaluate the utility of the proposed indicator in small-area disease risk assessment, the indicator was added to a set of area-level covariates in a restricted spatial regression that modeled the area-level risk of Type 2 Diabetes mellitus (T2D). The inclusion of the indicator improved model fit.

Objective 3: Based on store size and road network distance from origin (census dissemination area) to destination (store), I calculated the gravity-based pairwise probability of residents in each dissemination area visiting each store in the Census Metropolitan Area (CMA) of Montreal. Store-level sales were partitioned proportional to these visit probabilities and assigned to surrounding census Dissemination Areas (DAs). The resulting area-level purchasing indicator for soda, diet-soda, yogurt and plain yogurt improved the fit of T2D risk model, fit via a modified Besag-York-Mollié CAR model containing a scaled spatial random effect to enhance the interpretability of hyperpriors along with other area-level fixed effects.

Conclusions

My dissertation introduced and illustrated public health uses of POS data, highlighting their ability to improve nutrition surveillance, while developing and evaluating methodological approaches to address limitations of these data. My research thus forms an important foundation for further exploration and ultimately widespread application of these data in public health research and surveillance.

Résumé

Contexte

La mauvaise alimentation est le plus important facteur de risque évitable de la mortalité et de l'incapacité. L'alimentation peut varier à travers les villes et parmi les sous-divisions municipales en raison de la variation géographique des caractéristiques socioculturelles et économiques, du marketing et de l'environnement bâti. Le développement et l'évaluation efficace de politiques de nutrition et d'interventions dans la communauté dépendent de la disponibilité de mesures qui reflètent les différences géographiques en alimentation et les facteurs environnementales influençant les choix de nourriture. Une source traditionnelle de données pour la surveillance de la nutrition est les enquêtes sur la santé populationnelle, qui ont un échantillon inadéquat pour fournir des estimations de la nutrition de la communauté. De plus, les enquêtes ne peuvent pas mesurer le marketing important d'alimentation malsaine à l'intérieur des magasins telle que la réduction de prix, qui peuvent avoir des effets locaux et temporels. Les dossiers électroniques de transactions de point de vente d'épicerie (plutôt qu'au niveau de client individuel) qui sont générés par la lecture de produits ont le potentiel de résoudre ces limitations avec des approches analytiques appropriées.

Objectifs

Le but principal de cette thèse est le développement et l'application de mesures qui améliorent la capacité de la surveillance de la nutrition, en se concentrant spécifiquement sur un indicateur de santé publique du marketing de nourriture en magasin (Objectif 1) et un indicateur d'alimentation d'une haute résolution géographique (Objectifs 2 et 3) :

Objectif 1 : L'estimation de la modification de la mesure d'effet des rabais sur les achats de boissons gazeuses, en fonction du niveau d'éducation du quartier du magasin.

***Objectif 2*:** De développer un indicateur au niveau de petit domaine de la demande latente de boissons gazeuses en utilisant la modélisation hiérarchique bayésienne et d'en évaluer son utilité en estimant le risque du diabète de type 2 au niveau régional.

***Objectif 3*:** De développer un indicateur au niveau de petit domaine de boissons gazeuses, de boissons gazeuses diètes, de yogourt sucré et de yogourt nature en utilisant un modèle de gravité de magasins et d'en évaluer son utilité en estimant le risque du diabète de type 2 au niveau régional.

Méthodologie et Résultats

Objectif 1: J'ai modélisé les ventes agrégées de boisson gazeuses au niveau de magasins en fonction des rabais en magasin, des variables confusionnelles au niveau de magasins et de régions, et d'une interaction entre les rabais et le niveau régional d'éducation dans la région de tri d'acheminement où le magasin est situé. L'effet estimé des rabais sur les ventes de boissons gazeuses était plus grand pour les magasins dans les quartiers ayant un plus faible niveau d'éducation, surtout dans les pharmacies.

Objectif 2: J'ai modélisé la proportion des ventes de boissons gazeuses en magasin parmi les ventes totales de toutes boissons en fonction de covariables aux niveaux de magasins et de régions par une modélisation hiérarchique bayésienne avec un effet aléatoire spécifié par une distribution a priori autorégressive conditionnelle (ARC). La moyenne a posteriori estimée de la distribution a priori ARC est l'indicateur proposé de la demande latente de boissons gazeuses au niveau de petit domaine et peut être estimé pour toutes régions par lissage, incluant les régions qui n'avaient pas de magasins. Pour évaluer l'utilité de l'indicateur proposé dans l'évaluation de risque régional de maladie, l'indicateur a été ajouté à un ensemble de covariables régionales dans une régression spatiale restreinte qui modélise le risque régional de diabète de type 2. L'inclusion de l'indicateur a amélioré l'ajustement du modèle.

Objectif 3: Selon la taille de magasin et la distance routière entre le point d'origine (aire de diffusion du recensement) et la destination (magasin), j'ai calculé la probabilité par paires gravitationnelle que les résidents de chaque aire de diffusion aient visité chaque magasin dans la région métropolitaine de recensement. Le ventes au niveau de magasin ont été réparties proportionnellement aux probabilités de visites et assignées aux aires de diffusions comprenant les magasins. Les indicateurs correspondants représentant les ventes de boissons gazeuses, de boissons gazeuses diètes, de yogourt sucré et de yogourt nature ont amélioré l'ajustement du modèle de risque de diabète type 2. Cet ajustement a été fait par un modèle Besag-York-Mollié modifié ayant un effet aléatoire spatiale à l'échelle afin d'améliorer l'interprétabilité de l'hyperparamètre a priori avec d'autres effets fixes au niveau régional.

Conclusions

Ma thèse introduit et démontre l'utilité pour la santé publique des données au point de vente, en soulignant leurs bénéfices pour la surveillance de la nutrition, ainsi que des approches méthodologiques pour résoudre les limites des données. Ma recherche établit une fondation pour de futurs explorations et pour une application éventuellement répandue des données.

Statement of financial support

Acknowledgements

It is my great pleasure to acknowledge the support I received from my doctoral committee, lab mates, the department and my family. My thesis supervisor, Dr. David Buckeridge, provided an opportunity to tackle a highly unique research topic with a continuous support. Seeing his enthusiasm, creativity and insights in innovation-driven research was an invaluable experience. I also appreciate his flexibility and understanding for my frequent leaves due to family matters. I greatly thank my co-supervisor, Dr. Alexandra Schmidt, for her close guidance that allowed me to achieve the principled application of statistics to applied research. Aside from technical aspects, there was a great deal of learning from her devotion and commitment in teaching and research.

I am incredibly grateful to my thesis committee, Dr. Erica Moodie, who was instrumental in all aspects of my thesis work and provided moral encouragement in difficult times. I would like to express many thanks to Dr. Yu Ma for his detailed and timely responses to my endless questions about analysis, data and substantive knowledge in marketing science.

I also appreciate the members of the McGill Clinical and Health Informatics research group for providing close support for data extraction and management, analysis and technical writing. Equally importantly, I was saved by countless fun activities including soccer, daily darts and potlucks. I also thank Katherine Hayden, Andre-Yves Gagnon and Andrew Griffin in the Department of Epidemiology, Biostatistics and Occupational Health for providing administrative support and advices.

I would like to express my deepest gratitude to my family and relatives. I thank their support and understanding for my frequent absence in important family times for the last 6 years. I wish to thank my parents who supported me in all my pursuits and made me who I am today. You will be missed.

Statement of Financial Support

I would like to express my gratitude for the financial support I received during my doctoral training. Stipend was provided by the Department of Epidemiology, Biostatistics and Occupational Health (2013-2014) and by Dr. David Buckeridge (2016-2019). Scholarship was provided by the McGill University's Faculty of Medicine (2014-2016). My attendance to conferences was supported by a travel award provided by the Canadian Institute for Health Research (2016), the International Society for Disease Surveillance (2016), and the Department of Epidemiology, Biostatistics and Occupational Health (2015, 2016 and 2017).

Statement of Contribution

I (HM) formulated overarching research aim and specific research goals for all manuscripts in this thesis under the guidance of my primary supervisor (Dr. David L. Buckeridge), co-supervisor (Dr. Alexandra M. Schmidt), thesis committee members (Dr. Erica E. M. Moodie and Dr. Yu Ma). I also took the primary role in study design, data analysis, interpretation of the results, and preparation and revision of all manuscripts.

Dr. Buckeridge (DLB) oversaw the progress of my thesis and was involved in the research design, development of statistical analysis and interpretation of results, as well as providing financial support. Dr. Schmidt (AMS) joined my thesis committee after the completion of Manuscript 1. AMS supervised the design of statistical analysis and interpretation of the results for Manuscript 2 and 3 in collaboration with Dr. Moodie (EEMM). EEMM provided the main supervision for the design, analysis and interpretation of the findings in Manuscript 1. Dr. Yu Ma (YM) guided the preparation and management of grocery transaction data, including calculation of serving quantities, price discounting, categorization of food items into soda and other groups, for all Manuscripts. YM also provided substantive inputs and key references in consumer behavior for all manuscripts.

I drafted and revised all manuscripts in this thesis. They were critically reviewed and approved by all members of my thesis committee.

Manuscript 1: Susceptibility to Price Discounting of Soda by Neighbourhood Educational Status: An Ecological Analysis of Disparities in Soda Consumption Using Point-of-Purchase Transaction Data in Montreal, Canada.

Hiroshi Mamiya, Erica EM. Moodie, Yu Ma, David L. Buckeridge.

Published: *International Journal of Epidemiology*, Volume 47, Issue 6, December 2018, Pages 1877–1886, https://doi.org/10.1093/ije/dyy108.

Conceptualization and design: HM, reviewed by YM, EMM and DLB

Acquisition of data: DLB and YM

Analysis: HM guided by EMM

Interpretation: HM guided by EMM, YM and DLB

Manuscript preparation: HM

Critical review and approval of final version: EMM, YM, DLB

Manuscript 2: An Area-Level Indicator of Latent Soda Demand: Spatial Statistical Modeling of Grocery Store Transaction Data to Characterize the Nutritional Landscape in Montreal, Canada.

Hiroshi Mamiya, Alexandra M. Schmidt, Erica EM. Moodie, Yu Ma, David L. Buckeridge.

Published: *American Journal of Epidemiology*, Volume 188, Issue 9, September 2019, Pages 1713-1722, https://academic.oup.com/aje/issue/188/9#924158-5492131

Conceptualization and design: HM guided by AMS and EMM, reviewed by DLB and YM

Acquisition of data: DLB and YM

Analysis: HM guided by AMS and EMM

Interpretation: HM guided by AMS, EMM, YM, and DLB

Manuscript draft preparation: HM

Critical review and approval of final version: AMS, EMM, YM and DLB

Manuscript 3: Generating Area-Level Purchasing From Store-Level Sales: Application of a Retail Gravity Model to Create Small-Area Indicators of Food Purchasing in Montreal, Canada.

Hiroshi Mamiya, Alexandra M Schmidt, Erica E M Moodie, Yu Ma, David L Buckeridge

Conceptualization and design: HM, reviewed by AMS, YM, EMM and DLB

Data acquisition: DLB and YM

Analysis: HM guided by AMS and EMM

Interpretation: HM guided by AMS, EMM, YM and DLB

Manuscript draft preparation: HM

Critical review and approval of final version: AMS, EMM, YM and DLB

Statement of Originality

My doctoral research advances the science and practice of public health nutrition, with a focus on providing measurement capacities for within-store food marketing and heterogeneity of diets across small geographic areas. To achieve this goal, I demonstrated analytical methodologies adapted to the public health application of electronic Point-Of-Sales (POS) scanner transaction data.

Manuscript 1 is the first study modeling (aggregated) consumer behavior using POS data in the public health context. I investigated the presence of nutrition disparity created by price discounting, one of the most important promotional tactics of unhealthy food by retailers and food manufactures. Due to the prohibitive costs of surveying time-varying product pricing and purchase response for numerous food items in a store, temporal measurement of discounting has been largely inaccessible to public health researchers. Using the automatically recorded weekly fluctuation in pricing and sales for soda items in POS data, I conducted the first study estimating the association of discounting with soda sales from many stores, including supermarkets, pharmacies, supercenters, and convenience stores, across neighborhood education. The manuscript provides important evidence about the potential effectiveness of regulating price discounting, as the discounting of soda in pharmacies and convenience stores was shown to have a disproportionately higher association with soda sales in neighborhoods belonging to the lower tertile of education.

In manuscripts 2 and 3, I describe the development and evaluation of model-based methods for indirect estimation of population diets at a small spatial scale. Small-area indicators of diet are indispensable instruments for geographic surveillance of nutrition disparity across urban communities but cannot be estimated from nutrition surveys due to their limited sample sizes. The defining challenge of generating a small-area measurement of diets from point (store) sales is the absence of stores in some areas whose residents shop in neighboring areas. Two distinct and original approaches to solve the issue of 'borrowed sales from neighbours' are the smoothing property of Conditional Autoregressive (CAR) prior (Manuscript 2) and the partitioning and distribution of store-level sales into surrounding areas using the flow of residents predicted by a retail gravity model (Manuscript 3).

In Manuscript 2, I describe the first public health application of spatial statistical analysis to POS data in the form of a hierarchical spatial Bayesian model. Store-level sales of soda were modelled as a function of store and store neighborhood attributes in addition to a proper CAR prior representing latent area-level demand of soda. The use of a proper CAR prior allowed estimation of the spatially smoothed latent demand for soda, which was estimated even in areas that did not have stores. The proposed indicator of soda demand was the posterior mean of these random effects. In the process of estimating neighborhood-level relative risk of Type 2 Diabetes (T2D) prevalence using the proposed indicator as covariate, I also highlighted the shortcomings of conventionally used geographic food environment (i.e. "food desert") indicators as inappropriate surrogates of area-level diets to predict the risk of NCDs. Finally, the study introduced an approach to account for spatial confounding, an emerging issue in spatial statistics.

In manuscript 3, sales of four food groups (soda, diet-soda, plain yogurt, and flavored yogurt) in each store were allocated to surrounding areas, where the amount of sales allocated was defined by the pairwise interaction probability between destination (stores) and origin (areas) for residents' shopping trips. The probability was generated by the Huff gravity model, a utility model for store selection originally developed in retail marketing science to predict the flow of residents from residential areas to stores. Small-area measures of purchasing were then calculated as the sum of the partitioned sales from a set of surrounding stores.

The primary contribution of this thesis is a methodological foundation to enable the use of POS data to address existing research gaps in public health nutrition. Characterization of the strengths and development of research methods addressing challenges of these data will allow targeted and effective use of POS data to enhance measurement capability in public health nutrition research and surveillance, therefore advancing population health strategies aimed at prevention of nutrition-related chronic diseases.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIC – Akaike's Information Criterion

BMI – Body Mass Index

BYM–Besag-York-Mollié

CAR – Conditional Autoregressive

CBP – Canadian Business Point of interest

CCHS – Canadian Community Health Survey

CDC – U.S. Center for Disease Control and Prevention

CI – Credible (Confidence) Interval

CMA – Census Metropolitan Area

DA – Dissemination Area

DALY – Disability-Adjusted Life Year

DIC – Deviance Information Criterion

FSA – Forward Sortation Area

ICD – International Classification of Diseases

IQR – Interquartile Range

MCMC – Markov Chain Monte Carlo

mRFEI – modified Retail Food Environment Index

NAICS – North American Industry Classification System

NHS – National Household Survey

POS – Point-of-Sales

RSR – Restricted Spatial Regression

SAE – Small Area Estimation

SES – Socio-Economic Status

SGLMM – Spatial Generalized Linear Mixed Model

SIC – Standard Industrial Classification

SSB – Sugar Sweetened Beverage

T2D – Type 2 Diabetes mellitus

UPC – Universal Product Code

WAIC – Watanabe-Akaike Information Criterion

Chapter 1: Introduction

1.1 Background

Non-communicable diseases (NCDs) account for the majority of global premature deaths and disability, and unhealthy diet has become the leading preventable cause of NCD-related global premature death and disability in recent years (1–3). Diet-related disparity is the unequal distribution of dietary intake and behaviors among socio-culturally and economically marginalized countries and communities (4). Food preferences and selection are shaped by socio-cultural and community contexts and physical (e.g. greenspace) and retail environmental factors, including food marketing (5–7). Because these environmental factors are distributed unequally across urban neighborhoods or towns, they manifest in local disparities in healthy eating and NCD occurrences (8).

Interventions to minimize the prevalence of unhealthy diets, in particular among communities under the elevated risk of NCDs, are urgently needed to mitigate the societal and individual burden of NCDs (9). Policies and community interventions aimed at creating health-promoting environments have received greater attention in recent decades due to their sustainable and equitable effects on health (10–12). Successful development, evaluation and targeting of food environment interventions rely on knowledge of how environmental risk factors influence diets and on measures of dietary patterns across time, geography and socio-demographic and economic status (13,14).

Among the environmental risk factors, marketing of ultra-processed (unhealthy) food by food retailers and manufactures is one of the least investigated yet most powerful drivers of nutrition disparity (9). Product promotions within stores are recognized by marketing researchers as influencing purchasing decisions and are applied predominantly to promote ultra-processed foods (15,16). Because a manual in-store audit to capture (typically weekly) time-varying pricing is costly, the role of price discounting in driving unhealthy diets, especially among socio-economically disadvantaged individuals, is unknown (15,17). Therefore, regulating price discounting is an untapped and potentially effective policy option, although proposed interventions exist (18).

Another measusment gap in public health nutrition is the lack of an indicator that captures geographic variation in community diets. Mapping of disease events and health-related risk factors has long been a core function of population health assessment (19). At a fine spatial scale, information on community health status facilitates the development of local health interventions and the evaluation of community-specific responses to public health policies (20–22). Aetiologically, investigating the variation of local diets across socio-environmentally similar neighborhoods enables exploring unidentified barriers of healthy diets (23). Currently, estimates of consumption are available only at the level of large administrative areas (e.g. province) due to the limited sample size of national nutrition surveys in Canada and other countries (24,25). Thus, the capacity to systematically and routinely assess geographic disparity in nutrition has not been available to public health agencies and local health departments.

Point-Of-Sales (POS) data consist of product-level sales quantity and marketing attributes indexed with a temporal indicator of transaction (typically week) and the store where the transaction occurred. Commercial POS data are sold by global marketing companies in the form of a centralized transaction records, which are automatically and routinely collected from a cohort of sampled retail food outlets including supermarkets, supercenters, convenience stores and pharmacies across multiple retail chains (26). Originally collected to plan and evaluate product marketing strategies and track aggregated consumer behaviors over time and space, these data record temporal fluctuations in pricing and patterns of consumer food selection in store neighbourhoods (27). Appropriate analytical methods adapted to these data will provide an opportunity to overcome the measurement challenges described above.

1.2 Research Objectives

The goal of this thesis is to develop and evaluate methods for using POS data to measure within-store food marketing and community diets for small geographic area.

Although the methods developed could be applied to multiple types of food, soda (carbonated soft drinks) is the target food group of analysis in this thesis due to its role as the primary source of artificially added sugar. However, in Manuscript 3 I also consider other food categories, namely diet-soda, plain-yogurt and flavored yogurt.

Objective 1: To estimate the effect measure modification of price discounting by store neighbourhood education on soda purchasing.

Objective 2: To develop a small area-level indicator of latent soda demand using a hierarchical spatial Bayesian model and evaluate its usefulness in estimating area-level risk of type 2 diabetes mellitus.

Objective 3: To develop small area-level indicators of soda and yogurt sales using retail gravity model and evaluate their usefulness in estimating area-level risk of type 2 diabetes mellitus.

1.3 Format

This is a manuscript-based thesis composed of seven chapters. In chapter 1 (this chapter), I provide background and describe my research goal and three research objectives. In chapter 2, I contextualize the research aim through a literature review that summarizes the burden of nutrition-related chronic diseases, determinants of dietary behaviors, nutrition interventions, and the current state of dietary surveillance. In Chapter 3, I introduce the data and research design of the three objectives. In Chapters 4-6, I describe each objective as Manuscripts 1-3, respectively. Finally, in Chapter 7, I provide a summary of the findings and discuss the public health implications the results and future research opportunities. In Chapter 7, I also consider the novel aspects and limitations of the study, followed by a concluding section. Citations for all literature referred to in this dissertation are included in the bibliography at the end of the document.

Chapter 2: Literature Review

2.1 Nutrition-Related None-Communicable Chronic Diseases

NCDs are responsible for an estimated 70% (41 million) of global deaths, and 15 million of these are premature deaths occurring at ages less than 70 years (28). Among NCD, cardiovascular diseases and cancers are the leading cause of mortality, claiming 18 million and 9 million deaths in 2016, respectively (29). Type 2 Diabetes mellitus (T2D) is a major predisposing factor for cardiovascular diseases and other serious NCDs (30). T2D is rapidly becoming a major public health crisis in developing nations with economic and lifestyle transitions, while imposing significant societal and individual burdens on Canada and other developed nations (30–32).

The development of these NCDs are largely the consequences of behavioral (lifestyle) risk factors; notably unhealthy diet, tobacco smoking, lack of physical activity and excess alcohol consumption (2). Of these factors, a nutritionally inadequate diet is the leading preventable factor of global death and disability, responsible for an estimated 11 million deaths and 241 million Disability Adjusted Life Years (DALYs) in 2013 (33). The reported financial burden of NCDs due to unhealthily diet is estimated to lie between CAD $13.8 billion to $26.3 billion for direct health care costs, likely exceeding the burden of other behavioral risk factors, such as physical inactivity (34).

2.2 Unhealthy Diets

Unhealthy diet is defined as consumption patterns which deviate from national nutrition guidelines in terms of quantity and balance (35). Although 'healthiness' of diet varies by individual (e.g. age, gender, extent of daily physical activity) and cultural context, official nutrition guidelines generally define healthy diet as the optimal daily intake of healthy foods including fruits and vegetables, legumes, whole-grains, nuts and seeds. The definition also considers limited intake of free sugars, fats (in particular unsaturated and trans-fats) and sodium, typically through red or processed meats and ultra-processed foods (36). Ultra-processed foods are formulations of industrial ingredients and additives with little or no intact food, highly palatable and consumed for convenience due to their durable and ready-to-eat nature (37). These foods are associated with diets high in sugars, sodium and fats and low in protein, essential fibre, minerals and vitamins. These foods are typically sold as packaged items such as sugar-sweetened beverages (e.g. soda), snacks, chocolate and candies, and take-out meals (e.g. "fast food").

2.3 Obesity and Overweight

Obesity and overweight are the major manifestation of excess caloric intake (often coupled with a lack of physical activities); they are one of the most pressing public health issues of the 21st century (5). Obesity is a condition defined by having a Body Mass Index (BMI) of 30 kg/m^2 or higher, and overweight is defined as a BMI of 25.0 kg/m^2 or higher and less than 30 kg/m^2 (38). Obesity is a risk factor for many adverse clinical and physiological conditions, including hypertension, hyperlipidemia, osteoarthritis and other musculoskeletal disorders, cardiovascular diseases, and many forms of cancers (39).

Among NCDs, T2D is the most widely observed comorbidity of obesity and overweight: more than 87% of adult T2D cases were obese or overweight in the U.S, between 2011 and 2014 (40). Obesity also has debilitating consequences on children's health through psychosocial co-morbidities (e.g. stigmatisation, depression), asthma and limited physical functioning (41).

The global prevalence of obesity experienced nearly 10- and 3-fold increase over the past 3 decades among children and adults, respectively (42). In Canada, 1 in 10 children and a quarter of adults were classified as obese in 2013, and the prevalence was particularly high among individuals with a lower socio-economic status (SES) or First-Nation status (43–45). The observed increase in prevalence is largely attributed to societal and economic changes towards a sedentary lifestyle, transition of food system towards mass generation and distribution of affordable ultra-processed foods, and ubiquitous food marketing activities of these nutrition-poor energy-dense food (5,6). The increase in obesity prevalence is correlated with an increase of unhealthy food intake, a relationship that is especially pronounced in economically developing middle-income countries (5). Although the upward trend in obesity seems to have plateaued in many high-income nations (albeit at high levels), there is a remarkable *within*-country heterogeneity in obesity prevalence across geography and socio-economic classes (8,42).

2.4 Carbonated Soft Drinks and Yogurt as Source of Free Sugars

Excess intake of free sugars through ultra-processed food, most notably carbonated soft drinks (soda), is unequivocally linked to weight gain and obesity (46). Free sugars are all monosaccharides and disaccharides artificially added to foods by the manufacturer, cook or consumer, and sugars naturally present in honey, syrups and fruit juices (9). Products whose sugar is intrinsic (natural), such as fresh fruits and plain yogurt (lactose), are thus considered as devoid of free sugars. On the other hand, the sweet flavors of ultra-processed food products including soda are generally made up of free sugars only, while healthier products such as flavored yogurt also contain varying amounts of free sugar – often higher in quantity in products marketed to children. Estimated free sugar intake among Canadians in 2015 is 22 and 17% of total daily caloric intake for adolescents and adults, respectively (47), exceeding the recommended limit of 10% by the World Health Organization (48). In addition to weight gain and poor health outcomes including T2D (50), a higher intake of added sugar is linked to binge eating and reduced intake of essential micronutrients (50,51).

Sugar Sweetened Beverages (SSB) consist of any liquids containing free sugars including soda, sports and energy drinks, fruits drinks and sweetened milk, coffee, tea and water. Beverages with non-caloric sweeteners (e.g. aspartame, sucralose, xylitol) instead of free (caloric) sugars are not considered as belonging to the class of SSB and are labelled as diet or sugar-free products (52). Among SSB, carbonated soft drinks (soda) were the top source of free sugars, accounting for 7% and 9% of daily total sugar intake among adolescents and adults, respectively, in Canada in 2015 (53). Consumption of SSB has an epidemiologically established association with increased BMI and T2D (46) and has a link with cardiovascular diseases, some cancers, kidney diseases, tooth decay and cavities. Globally, SSB consumption is responsible for an estimated 184,000 (95% CI: 161,000 - 208,000) deaths, of which the majority occur in middle and high-income countries (54).

Although perceived to be healthy and a less important caloric source than SSB, yogurt can be a hidden source of sugar (55). Yogurt is defined as a fermented milk product containing digested lactose (56). The marketing and consumption of these products in North America experienced a rapid expansion in recent years due to functional claims for constituents such as probiotics and calcium and acceptability to lactose tolerant individuals (57–59). Although epidemiologic and biological research generally supports its protective effects against weight gain, T2D, and cardiovascular diseases (56,60,61), yogurt items available in the market are over-represented by artificially sweetened products with a varying amount of free sugars (in addition to naturally occurring lactose) to enhance texture and flavor (62). Sugar-sweetened yogurt therefore has the potential to lead to diets high in free sugar, especially items marketed to children that have been shown to have a higher quantity of free sugar than products marketed to adults (63).

2.5 Global and Canadian Trends of Diet

Although dietary trends vary radically across countries, most nations experienced an increased intake of healthy food between 1990 and 2010, likely as a result of technological enablers for automated and large-scale food production, distribution and storage (64). Unfortunately, mass production, marketing and global trading of ultra- processed foods also caused the consumption of unhealthy food to outpace that of healthy food during the same time period in most areas, in particular among middle-income nations (6,64,65). Many lowest-income nations did not experience an increase in healthy food consumption due to limited access to healthier foods among the poor, while still having an over-abundance of unhealthy foods. These nations thus suffer from obesity and related NCDs as well as stunting, cognitive underdevelopment and immunodeficiency due to the lack of essential nutrients (65). Finally, although high-income nations experienced a reduction of unhealthy food consumption, the downward trend is likely to be true only among socio-economically well-off populations, leaving the nutritional gap unaddressed (8). Taken together, it is clear that achieving optimal population diets remains a major challenge for all nations.

The diets of Canadians showed a modest improvement across two cycles of a national survey, the Canadian Community Health Survey (CCHS) 2004 and 2015 when measured via the Healthy Eating Index score, but the majority of Canadians still failed to follow national dietary guideline (37,66). The consumption of fresh vegetables and fruit and milk was lower in 2015 than in 2004, although there was a shift towards plant-based protein rather than meat, as recommended (67). Although there was a decline in the consumption of SSB, these beverages remained a major source of daily energy intake, especially among youth and population subgroups at risk of NCD (37) . Soda remained the top source of free sugar among SSB, and one of the major sources of total sugar along with healthy alternative sources of sugars, milk and fresh fruits. The overall caloric intake of free sugar was not meaningfully lower in 2015 compared to 2004 for adults or children, once misreporting of dietary recall was taken into account (53).

2.6 Individual and Household Determinants of Diet

Socio-economic status (SES), as commonly measured by education and income, is shown to exert considerable influence on diet and other behavioral risk factors (68). As a mediator for various determinants of health, SES is also widely used as a key indicator of health inequalities (69). Household or personal income is a direct measure of financial and physical opportunities (e.g. housing location and private transportation) to obtain healthy foods that are typically more costly than ultra-processed foods per caloric base (68). A high income thus correlates with food security (70) but does not necessarily lead to a lower consumption of unhealthy food, perhaps because individuals, communities and countries with disposable income can afford more of both healthy and unhealthy foods (71). On the other hand, empirical investigation of SES-specific diets suggests that education is more consistently positively correlated with diet than income (72–75). The positive benefit of education on health-promoting behavior is rooted in better self-management, better awareness of elevated disease risk from unhealthy diets, and higher potential future occupational status and earning (69).

In addition to SES, dietary patterns are known to vary across demographic attributes including gender and ethnic and cultural background (64). Females and recent immigrants in particular tend to have healthier diets and a lower risk of NCD, the latter being due to the selection effects of immigrants though medical screening (76). These factors are associated with many intertwined determinants of dietary preference and selection including values, motivation, habits, financial capacity, skills and knowledge in food preparation, and susceptibility to marketing exposure (68,77).

2.7 Food Environment and Marketing

There is a consistent finding that the heterogeneity of health behaviors across individuals is not fully explained by personal characteristics (e.g. household education and income) but rather is attributed in part to the environmental conditions in which people live (78,79). As an example, built environmental features, such as proximity to recreational facilities and supermarkets, have been widely used to investigate a city's physical influence on obesogenic behaviors (80,81). These environmental features are distributed unequally in space and are associated with diets, independently or jointly with the aforementioned individual-level factors (5,82).

A substantial part of the obesity-inducing environment is formed by the mix of food marketing activities by retailers and manufacturers (6,10,16,83). These promotional activities affect food preferences and behaviors at all stages of a consumer's life, shaping childhood food preference by television advertisement, influencing where people shop through store location and physical accessibility (e.g. parking), attracting a wider range of consumers by increasing product assortment (e.g. diversity of flavors), and inducing planned or impulsive purchasing through discounting (6,84,85). As a modifiable environmental target of regulatory and policy interventions, these marketing activities have been under increasing research attention (86,87).

Food marketing activities are broadly classified into three categories: Spatial measures of store location within one's neighbourhood as "*community food (nutrition) environment*": within-store food marketing as "*consumer food environment*", and food advertising activities outside stores e.g. television advertisement as "*information environment"* (7).

2.7.1 Community Food Environment

The availability and types of food outlets in neighborhoods has been hypothesized to affect the dietary quality of residents (79). The community food environment represents the geographic store availability and is typically measured as the count or density of food outlets categorized as healthy or unhealthy, with their location and store type determined by commercial or government business point-of-interest databases (88). In this binary scheme, stores carrying fresh fruits and vegetables, primarily supermarkets, are labelled as 'healthy', whereas outlets without these food (e.g. convenience stores) are deemed 'unhealthy', as the choices offered by the latter store type are limited to highly processed food (90). To generate a map of obesogenic environments, residential areas can then be classified as *food deserts* (areas lacking healthy stores), or *food swamps* (areas overpopulated by unhealthy stores). One of the most widely used indicators of community food environment is the U.S Center for Disease Control and Prevention modifiable Retail Food Environment Index (mRFEI), which is calculated as the ratio of the number of healthy stores and all stores in a given area (89). Food environment unsupportive of healthy diets is therefore defined as areas with a zero value of mRFEI (food desert) and a smaller (non-zero) values of mRFEI (food swamps).

Community food environment measures are far more widely available than consumer (in-store) environmental measures because of the widespread availability of business location data (90). Community food environment measures can help to identify inequalities in the retail built environment (88,89,91–93) but these measures do not necessarily predict local diets and NCDs. Despite many studies, systematic reviews have failed to note an association between spatial store access, diet and NCDs, with the exception of some studies conducted in the U.S. (94,95).

The inconsistent association between community food environment and NCDs may be attributable to unmeasured yet influential area-level factors of diets, most importantly food price. Because a food desert is formally defined as the lack of access to *affordable* food (96), there is a need to measure neighborhood food pricing for a range of (un)healthful items. Further, although defined as "healthy" store type, supermarkets are a primary supplier of unhealthy as well as healthy foods (97). Therefore, although areas with an abundance of supermarkets and no unhealthy stores receive a favorable score for healthy eating (i.e. not classified as a food desert or swamp), these areas have a high spatial accessibility to unhealthy food. Furthermore, geographical access measures face methodological limits as the areal unit of analysis becomes smaller. This is because numerous areas will have no stores at all (a zero count, i.e. a complete food desert), even though shopping activities of residents in these areas take place in neighboring areas. Methodologically, the overwhelming majority of food environment studies in the past three decades used multi-level analysis or ecological regression without inspecting or addressing the spatially correlated nature of health outcomes (98). The assumption of independence in these studies may have distorted the evidence base formed through food environment studies, specifically with biased point estimates and overly optimistic standard errors.

2.7.2 Consumer Food Environment and Price Discounting

Consumer food environment is measured by food marketing activities in retail settings and includes price, price discounting, assortment (diversity of products), couponing, display promotion (placement in prominent locations) and flyer promotion (15,16,18,99,100). This environment is usually measured through ad-hoc surveys of a sample of food items in each store, such as within-store audits to measure product-level promotions, which provides a reliable measure, but requires considerable resources (17,85). For this reason, research investigating the association of in-store food environment with diets or NCD risk is sparse and typically cross-sectional (101), even though many promotional activities vary by week.

Not surprisingly, price reduction appears to be the most effective in-store marketing tactic (16) and was shown to disproportionately target ultra-processed food (15). Price discounting can induce impulsive purchasing (84,102) and it is thought to hinder the effect of taxation policies aimed at discouraging the purchase of unhealthy products (103). Although gaining increasing attention as a potentially modifiable driver of nutrition disparities (15,18), only three studies as of 2019 attempted to evaluate the SES-specific effect of exposure to price discounting on unhealthy food purchasing using a representative sample of consumers (17,85,104).

2.8 Nutrition Inequality

The burden of adverse health status is positively correlated with a downward gradient of socio-economic status within countries at all levels of development (105,106). For example, the burden of obesity and food insecurity is higher in both low-income individuals and low-income communities (107,108). In addition, adverse health behaviors, such as unhealthy eating, lack of physical exercise and cigarette smoking exhibit clustering among socially disadvantaged individuals (75). The disparity arises from an interplay of individual and environmental factors, where the lack of personal or household resources to improve and sustain healthy behaviors negatively interacts with the lack of environmental opportunities for improving healthy lifestyle. As an example, the effect of shopping in a discount store (presumably more intense marketing of ultra-processed food) on unhealthy food purchasing was found to be stronger among mothers with a lower educational attainment (109). As a driver for obesity and T2D disparity, segregation of low SES communities from recreational facilities and green spaces was also reported (110). Other nutrition-damaging environmental exposures associated with socio-economically disadvantaged communities include inferior municipal health and non-health services, school quality, and job opportunities (111).

2.9 Food-based Nutrition Interventions

The goal of nutrition intervention is to provide safe, nutritious, affordable and sustainable diets to all members of a society. The unequal concentration of premature death and health-damaging behaviors among socially disadvantaged individuals underlines the urgency of preventive interventions for NCDs. Because those at the highest risk of NCDs tend to be doubly exposed to low individual (household) resources and an unfavorable environment for healthy lifestyle, an equitable population health strategy should include person- and place-(context) based intervention by government and community organizations (91,111).

2.9.1 People-Based Interventions

Based on principles from health education and physiology, numerous health promotion programs to build knowledge and awareness for healthy diets were delivered in specific setting such as hospital, school, workplace or disseminated publicly through mass media (112). Without accompanying policies to address environmental determinates of health, these educational programs have had a limited adoption by low-SES populations, resulting in increased nutrition disparity (77). Nevertheless, with a targeted use as part of community-rooted interventions, educational programs are an essential component of public health tools to modify behavioral risk factors among disadvantaged population subgroups (113).

2.9.2 Interventions Targeting Food Environment

The past three decades has seen an explosive growth in the number of experimental and observational studies that aim to identify the independent influence of a change in living context on diet and NCDs (79). The momentum was driven by inquiries for sustainable and equitable improvement of diets, which were hypothesized to be better achieved by providing environmental opportunities for healthy lifestyle (114–117). Aside from social and cultural aspects of neighborhood, the food marketing (consumer and community) environment includes many potentially modifiable factors that influence health (10). These factors have been the subject of extensive investigation and review (118).

Price-based interventions are an effective means of creating favorable changes to the retail food environment. One of the most widely proposed and investigated policy interventions is the taxation of ultra-processed, high-fat, or high-sodium food, an approach that has shown a promising effect on the purchasing of sugar sweetened beverages (119,120). The beneficial effect of taxation is likely to be reinforced by the simultaneous implementation of subsidization or discounting of healthy foods (121). Among price-based interventions, strategies to regulate the price discounting of ultra-processed food has received little policy and research attentions to date (18). A cross-sectional analysis of shoppers' baskets indicates that a greater proportion of purchases made by low and middle-income families were price discounted (104). Because of the potential effect of discounting to create accelerated consumption of ultra-processed food among at-risk populations, there is an urgent need for research estimating the impact of discounting across a range of consumers (18).

Other interventions to improve retail consumer food environments include display of healthy food items in prominent locations to increase awareness, voluntary nutrition reformulation (reduced sugar contents) by manufacturers, mandated changes in sugar, fat, and sodium contents in highly processed food, and easy-to-read nutrition labelling and warning on processed food (122).

In terms of improving the community food environment (physical accessibility to healthy and unhealthy stores), natural experiments of opening a supermarket in a food desert area have not resulted in dietary benefits, while observational methodologies to identify causal effects of neighborhood store mix on health-related behaviors are still under development (79). However, preventing food swamps in school zones by banning the opening of fast food restaurants appears to warrant further investigation (118).

2.9.3 Community Interventions

Public policies and regulations applied at a large geographic scale do not necessarily address geographically varying social, cultural and environmental barriers to healthy eating (22). National-level programs appear to be more effective when supplemented with local initiatives in the form of educational campaigns, health promotion, and changes to local food environments (123,124). As a target of interventions, a community may be defined as a town or the smallest level of city and metropolitan subdivisions at which local health, social, and other administrative services operate.

The effectiveness of community-affiliated actions is supported by the widely accepted socio-ecological approach to obesity prevention, whereby behavioral modification may be best achieved by concerted interventions addressing determinants rooted at multiple levels of influence e.g. family, socio-cultural and store environment, and provincial/national context (125,126). The major advantage of community interventions is a better identification of local context and tailored solutions that can be achieved in collaboration with community knowledge partners such as schools, retailers, care providers, and recreational facilities (124). Specific community programs could include independent or coordinated actions of school education and meal programs, change of food environment (e.g. providing fresh water supply, removing vending machines for SSB), educational booths at public events, online and media campaigns, removal of advertisement from public spaces, or a mobile food store or community market with adequate affordability (127–130).

2.10 Geographic Surveillance of Population Diets

Surveillance is a cornerstone of public health practice and decision making, as it provides information necessary for identifying vulnerable subpopulations, detecting emerging problems, prioritizing programs, establishing baseline health status to evaluate socio-economic and epidemic events, track and evaluate population-specific response to interventions, and generating hypotheses for etiologic investigations (13,131). To be informative, routine data collection and analysis followed by timely dissemination to decision makers and public is required. From a measurement perspective, comprehensive monitoring of population diets implies ongoing assessment of spatio-temporal trends in diets and environmental factors driving food-related behaviors (14).

To date, collection of dietary data has relied largely on questionnaires of dietary recall administered through a population health survey, such as the Canadian Community Health Survey (CCHS) nutrition module or the U.S. National Health and Nutrition Examination Surveys (NHANES). These data are cross-sectional and available across socio-demographic strata (e.g. age, education, income) at the provincial (state) level, but fail to provide adequate statistical precision at a small spatial scale due to the limitation of sample size. Methods for Small Area Estimation (SAE) of chronic diseases and their risk factors exist (20,21,132), but they typically require combining multiple years of moderately large surveys with the trade-off of supressing the temporal patterns of diets. The application of existing SAE methods is thus limited to the data from annually conducted general (not nutrition-specific) health surveys, which lack detailed dietary questionnaires. Because nutrition-specific surveys are administered very infrequently, for example once in every 10 years in Canada (43), they cannot be combined across survey cycles.

Consequently, currently available geographic indicators of nutrition in Canada and many other countries are limited to the level of province or large metropolitan area. The unavailability of local information on diet impedes the development and evaluation of community programs (106,130). Furthermore, public health agencies lack the means to assess across-community heterogeneity of dietary risk factors, information which is invaluable for resource allocation and exploration of unidentified environmental risk factors driving nutrition disparities. Mapping of health-related attributes is an integral component of public health surveillance, and electronic administrative health records and census surveys are routinely used to generate the community atlas of disease occurrence (prevalence or incidence) and social determinants of health, respectively. Despite being a strong predictor of NCDs, dietary activities are not estimated at the same spatial scale, and therefore are not available as a covariate to execute some of the fundamental tasks required for epidemiological or public health investigation, including disease risk estimation (commonly known as disease mapping), spatial regression, and forecasting of future disease burden in communities.

A costly and non-sustainable solution for this measurement gap is ad-hoc data collection as part of evaluation research for community interventions (127). The cost of such surveys can exceed the cost of the intervention itself when conducting detailed dietary assessment for a representative sample of participants (133). For this reason, environmental indicators of dietary response, such as changes in sales or even retail response (change in product selection in stores), have been proposed as important new tools for tracking and evaluating community programs (133–135).

2.11 Electronic Grocery Transaction Data

The electronic footprints of consumers, commercial industries and health systems allows efficient and timely collection of disease outcomes and lifestyle risk factors and environmental attributes that affect disease risks. Although these digital data play an increasingly important role in etiologic research and public health surveillance (136), the value of their contribution rests on appropriate research design and interpretation accounting for the secondary nature of these data (e.g. non-random sampling, data not collected by questionnaire). Research on the use of digital sales data to capture behavioral risk factors, including diet, is just beginning, but shows considerable promise given that these behavioral factors are some of the most poorly captured risk factor of chronic diseases (106).

The introduction of the Universal Product Code (UPC) and barcode scanning technologies in the 1970s led to the automated collection of consumer transactions from retail stores, largely replacing traditional sources of sales data generated by in-store audits, purchasing diaries, and surveys (27). Point-of Sales data are time and location-indexed transaction records in the form of item-level sales aggregated across customers for each retail outlet. In addition to sales quantity, the data contain weekly varying promotional activities associated with each item, for example pricing, display and flyer promotions.

Marketing research firms, such as the Nielsen company, collect POS data from multiple chain stores and store types (e.g. supermarket, pharmacy, supercenter, convenience store) in many countries (53). In marketing and retail management science, POS data have been used to investigate spatial and temporal trends of product sales, evaluate the effect of product marketing and recall, and the survival of new products (137,138). Automated product scanning allows researchers to measure the dynamics of sales and the consumer food environment (i.e. product marketing) across many items in sampled stores over a large geographic area. From a geographical perspective, because store-level sales are geocoded and represent aggregated consumer behaviors in surrounding areas, appropriate analytical methodologies should allow the estimation of dietary status of communities.

2.12 Summary

Diet is the most important modifiable risk factor for NCDs. Given the lower quality of diet and disproportionately higher burden of NCDs among socio-economically disadvantaged populations, interventions to minimize nutrition disparity are needed urgently. Equitable interventions should target modifiable components of the food environment that vary geographically and exert community-level variation of dietary patterns.

Food marketing within stores, or the consumer food environment, is a largely unevaluated environment for interventions due to the challenge of measuring time-varying promotional activities for many food items. Of these activities, price discounting of unhealthy food is suspected to play an influential role in driving purchasing of ultra-processed food products. Unfortunately, few studies have investigated the association of price discounting and food purchasing, and those that exist are cross-sectional in design. Longitudinal investigation of purchasing responses to price discounting across varying community socio-economic strata may motivate new policies or policy research in regulating discounting. Such study, however, requires the availability of data that captures the temporal dynamics of discounting.

Lack of dietary indicators at a small spatial scale is another measurement gap in public health nutrition. Although community-level interventions play a vital role in addressing locally varying socio-cultural and environmental barriers to healthy eating, development and evaluation of these programs and identification of high-risk communities cannot be achieved without indicators of diet being available at the geographical level of the community. Currently, estimates of diet are limited to the level of province due to the universally limited sample sizes of national nutrition surveys, which fail to capture important variation in dietary patterns at a scale that is meaningful for community-based initiatives.

Originally collected for marketing research, grocery POS data contain time-varying price discounting and sales response for all food products purchased at a geographically representative sample of retail food outlets. Because sampled stores are geo-coded, the data also provide a spatially-varying record of product sales, which may be used to generate community measures of food purchasing. Therefore, appropriate analytical methodologies to address secondary nature of these data should shed insights into the influence of price discounting on nutrition disparity and geographic distribution of dietary patterns at a fine geographic scale.

Chapter 3: Data and Methods

3.1 Study Design and Time Window

The three manuscripts are ecological in design, with aggregated product sales by store-week as the unit of analysis. Specifically, Manuscript 1 is a longitudinal analysis of store-level weekly soda sales generated between 2008 and 2012, inclusive. Manuscripts 2 and 3 are cross-sectional analyses, focusing on store-level annual sales and area-level T2D prevalence in year 2012. The analysis in Manuscript 2 followed two steps: first, store-level annual sales of soda were modeled (sales model) to generate the area-level indicator of latent soda demand, followed by estimation of area-level relative risk of T2D (disease mapping) using the proposed indicator as one of the covariates. The analysis in the third manuscript consisted of three steps: a sales model was fit as in Manuscript 2 for four food categories, then (predicted or observed) store-level sales were decomposed into area-level purchasing using a retail gravity model, and finally disease mapping of T2D was performed using the area-level purchasing indicators as covariates.

3.2 Geographic Window

The target geographic area on which I have focused in my thesis is the CMA of Montreal, a metropolitan area including the Island of Montreal and the North and South shores (off island), encompassing a population of 3,824,221 inhabitants as of 2011 (139). A CMA is one of Statistics Canada's standard geographical classifications, and its boundary is formed by adjacent census subdivisions that have sufficient interaction with each other in terms of commuting movement (140). Therefore, I expect that shopping trips originating in the CMA of Montreal largely end in stores located within the CMA.

3.3 Definition of Area and Areal Attributes

The study region (CMA of Montreal) was partitioned into non-overlapping small areas. The small areas represent store catchments (Manuscript 1 and sales models in Manuscripts 2 and 3), for which SES and demographic attributes were calculated. The areas also represent the spatial unit at which disease risk is assessed (disease mapping, Manuscript 2 and 3). Area-level attributes from these spatial units were available from the 2011 Canadian census and National Household Survey (NHS).

3.3.1 Forward Sortation Area

In estimating the association between store-level price discounting of soda and soda sales (Manuscript 1), area-level education (effect measure modifier) and area-level confounders were measured at the level of the Forward Sortation Area (FSA). This is a geographical unit defined by the first three characters of the six-character Canadian postal code and used in previous Canadian studies (141,142). On average, a single FSA contained 8000 households in 2011 (143). There were 200 FSAs in the CMA of Montreal in 2011.

3.3.2 Neighbourhood

Area-level covariates of sales models and disease mapping were measured at the level of Montreal neighbourhood in Manuscripts 2 and 3. Boundaries of neighborhoods were defined by manually merging contiguous census tracts to maximize the homogeneity of SES and demographic attributes (144). These areas were created by the public health department of Montreal, in collaboration with local community partners (144). As with the FSA, the neighbourhood is a small area (median, 5.72 km^2, Interquartile Range [IQR], 2.52–18.23) with a median population of 18,410 residents (IQR, 11,870–25,530 residents). There were 195 neighborhoods in the CMA of Montreal.

3.3.3 Dissemination Area

The Dissemination Area (DA) was used in Manuscript 3 to calculate origin-destination store visit probability. This is the smallest Canadian standard geographic unit in which census data are reported, with a population size ranging between 400 and 700 residents (145). At the time of the 2011 NHS survey, there were 6,235 DAs in the CMA of Montreal.

3.3.4 Area-level Socio-Economic and Demographic Attributes

The area-level measures used in this thesis are listed below (Table 3-1). SES and demographic measures were obtained from the 2011 NHS survey, while built environment indicators (recreational facilities and mRFEI) were obtained from the Canadian Business Point (CBP) of interest data discussed below.

Table 3-1 List of Covariates Likely to Correlate With Residential Location and Affect Obesity Measured by 2011 NHS survey, CMA of Montreal

Variable name	**Description**
Income[a]	Median household income in Canadian Dollars
Education[a]	Proportion of residents with post-graduate diploma, certificate or higher among those 25 years and older
Family size[a]	Average number of people per household
Population density[a]	Number of residents per square kilometer
Employment[a]	Proportion of employed residents among labour force
Age[a]	Proportion of residents under 18 years old
Percent immigrants[a]	Proportion of immigrants arrived within less than 5
Recreational facilities[b]	Number of recreational facilities per resident
mRFEI[b]	Number of healthy retail food outlets divided by number all (healthy and unhealthy) retail food outlets.

[a] Data obtained from 2011 NHS survey
[b] Data obtained from CBP

3.4 Point-Of-Sales Transaction Data

We purchased commercial POS data (tradename: MarketTrack) from the Nielsen company (26). These data are collected globally from affiliated retail chain stores through direct data transfer of electronic transaction records and store attributes (e.g. address) from participating retail chains and stores. As well, the Nielsen company performs periodic manual in-store audit to assess data quality, including geocoding of sampled stores.

3.4.1 Characteristics of Store Cohort

The Nielsen Corporation selects *chain* retail stores from the Montreal CMA by stratified random sampling, where strata are defined by urban/rural status, store size, and store type. The sampling frame includes chain retail outlets from the following store types: supermarket, pharmacy, and supercenter (e.g.. Walmart). Additionally, chain convenience stores and gas stations are sampled separately, again stratified by urban/suburban status, store size, and store type (i.e. convenience store or gas station). Stores not scanning transactions, independent (non-chain) stores, the warehouse Costco, and dollar stores are excluded from the sample frame. To maintain the representativeness of the sample in a continuously changing store mix, periodic partial resampling is performed. Thus, not all stores were followed up for the entire 6-year period.

Based on our preliminary work using CBP, the Neilson sampling frame represents 65.2 percent of the market share of all food establishments (chain and non-chain stores selling food items**)** in the Montreal CMA.

3.4.2 Description of POS Transaction Data

The Nielsen POS data consist of the store-weekly aggregated quantity of sales and information on each food product sold. Food products were uniquely defined by product name and the Universal Product Code (UPC) and labelled with product category (e.g. soda, yogurt). Time-invariant variables are the chain and store identification code in which a product was sold, the postal code of each store, as well as the product name and category. Retailers normally schedule promotional activities by week, thus time-varying product attributes and aggregated (sum) sales are available weekly. Weekly-varying variables are price, quantity, date, and a binary indicator for non-price promotional activities (flyer and display promotion). Display promotion is a form of marketing where products are placed at a visible location, such as the end of a shopping aisle and the store entrance to increase awareness of promoted products. Flyer promotion indicates that a product was included in flyers in a given week. Price indicates selling price for that week (after discounting if present), from which the reference price and percent discount was calculated.

3.4.3 Food Transactions and Sample Size

The target food category was soda in Manuscripts 1 and 2, whereas the focus of Manuscript 3 was expanded to include soda, diet-soda, plain yogurt, and flavoured (sugary) yogurt. For typical supermarkets, each of the four categories contains numerous products. Before aggregation into the category, the quantity of each food product sold was standardized using the Food and Drug Administration (FDA) standard serving size, which is 240ml for beverages and 170g (6-ounce) for yogurt.

The unit of analysis in Objective 1 was store-week, and observations consists of the following numbers of store-weeks sampled over the six-year interval: 18,743 supermarket-weeks (for 91 supermarkets), 12,437 pharmacy-weeks (for 57 pharmacies), 3,965 supercenter-weeks (for 42 supercenters), and 49,533 store-weeks (for 200 convenience and gas stores).

For the sales model in Manuscript 2, the average annual proportion of soda sales out of the total sales of all beverages from 125 sampled stores (60 supermarkets, 40 pharmacies, and 23 supercenters) in 2012 was used. These stores consisted of 6 supermarket chains, 5 pharmacy chains and 2 supercenter chains, all selling soda and diet-soda products. Note that there were 972 out-of-sample stores (i.e. stores not sampled by the Nielsen company, thus without observed sales) in 2012 as recorded in CBP. The missing sales in out-of-sampled stores was treated as the parameter to estimate in the soda sales model. For disease mapping, the number of areas (neighbourhoods) was 193 in the CMA of Montreal, after excluding two areas that belonged to federally registered First Nation communities. As the exact location of convenience and gas stores was difficult to determine, we excluded sales from these store types from the spatial analysis.

For objective 3, sales from the same store population in Manuscript 2 were used to model the sales of soda and diet soda. For yogurt, average annual natural-log transformed weekly sales of 60 supermarkets and 12 supercenters were used, excluding all pharmacies and one supercenter chain that did not sell yogurt. There were 239 out-of-sample stores (all supermarket chains and one supercenter chain) as indicated by the CBP, and the sales of yogurt in these stores, as for soda, were imputed by the mean of the posterior predictive distribution.

3.5 Prevalence of Type 2 Diabetes Mellitus

The outcome of disease mapping (Manuscripts 2 and 3) was the prevalence of T2D at the level of neighborhood in 2012. The cases of T2D were ascertained from a random sample of 25% of individuals from the Regie de l'Assurance Maladie du Quebec, the representative public health insurance registry covering 99% of the population in Quebec. The cases were identified in accordance with the T2D definition used by the Canadian Chronic Disease Surveillance System (146). Specifically, individuals receiving at least 1 hospital diagnostic code (*International Classification of Diseases [ICD], Tenth Revision*) of T2D or at least 2 diagnoses of T2D in the physician claims database (*ICD, Ninth Revision*) within a 2-year period were defined as cases of T2D. Exclusion criteria were gestational diabetes or being under the age of 2 years for both cases and non-cases.

3.6 Business Point of Interest Data

While sales transaction, store location and chain identification code of sampled stores were available in POS data, information on out-of-sample stores was available in the Canadian Business Point data. The CBP is an annually updated enumeration of businesses performed by a commercial company, Pitney Bowes Canada; it covers all operating and closed business establishments, including chain and independent stores and restaurants in Canada (147). The data include point store location (street address), store type classified by the North American Industry Classification code and Standard Industry Classification code, business type, business name, and business size represented by gross annual sales and the number of employees.

The physical presence of recorded supermarkets, convenience stores, and fast food restaurants in the CBP data were inspected by field audits previously, with a strong correlation with the actual establishment of these stores (148). Because the CBP also contains business information of recreational facilities, it provided the measure of area-level accessibility to recreational facilities per capita, which was used as a covariate in disease mapping of T2D (Manuscript 2 and 3). Another covariate representing food-related built-environmental feature in disease mapping, mRFEI, was also generated from the CBP data (Manuscript 2).

3.7 Bayesian Disease Mapping and Spatial Regression

Disease mapping is a model-based estimation of area-level disease risk, typically using widely available case-event data aggregated at the level of non-overlapping geographic areas. This approach is used to uncover the variation of disease risk across areas, and it can provide a means to rank communities and prioritize interventions (149). Disease mapping may also help to identify unrecognized environmental drivers of disease risk (150).

Analysis of survey data for small areas leads to sparse observations, resulting in estimated local risk driven by the random component of survey sampling and the rarity of diseases rather than reflecting the true risk. Measurements with spatial information also exhibit autocorrelation, where observations closer in space are more similar in value than those located further apart (in case of positive correlation) due to spatially structured unmeasured factors affecting disease risk. Failure to acknowledge the residual spatial variation will therefore result in spurious precision and may bias estimates of the risk. Autocorrelation and sparseness of observation are commonly addressed with spatial smoothing via spatially correlated random effects.

3.7.1 Spatial Generalized Linear Mixed Model

Assuming Poisson likelihood for case event data, the mean number of disease cases (area-level count) is a product of area-level relative disease risk and heterogeneous background occurrence of the disease. The latter is the area-specific expected count of the disease calculated by indirect standardization with a suitable stratum-specific reference rate, typically age and sex. The main quantity of interest, disease risk, is a log-linear combination of intercept, area-level covariates (if measured) as fixed effects and an area-level random effect. The random effect accommodates, *a priori*, residual spatial structure through a Conditional Autoregressive (CAR) prior distribution (151), where the prior distribution of each local random effect depends on the neighboring random effects. Because extra-

Poisson variation of the relative risk can be partially attributed to a spatially uncorrelated latent effect, it is often sensible to include an independent (non-spatial) normally distributed random effect, as suggested by Besag, York and Mollié (hence forming what is known as the BYM model) (152).

3.7.2 Spatial Confounding

Oftentimes research inquiry is on spatial regression: inference on the association of etiologic factor(s) and disease risk, rather than (or in addition to) the estimation of spatially smoothed disease risk. In a spatial generalized linear mixed model, covariates can correlate with spatially structured unmeasured confounders, and the use of a CAR prior was shown to result in a large change in the posterior distribution of the regression coefficients (153), a phenomenon known as spatial confounding or confounding by location (154,155). A proposed reparametrization of spatial regression, Restricted Spatial Regression (RSR), removes the confounding by restricting spatial random effects to residual subspace orthogonal to covariates (154,156). More recently, Hughes and Haran demonstrated a computationally efficient RSR, sparse SGLMM , where the geometry of projected spatial random effects orthogonal to covariates now include the underlying spatial structure (termed as Morans operator) (156). Eigendecomposition of this operator effectively reduces the dimension of the random effects by removing the patterns corresponding to negative spatial dependence (i.e. spatial repulsion) that is highly unexpected to occur in disease mapping.

3.7.3 Hyperparameters in the Besag-York-Mollie Model

In the BYM model, the variance of the spatially structured random effect is interpreted conditional on the effect from neighbors, while that of non-structured random effect represents the marginal variance of the independent effect (157). Therefore, these two variances are in different scales. An appropriate specification of the prior distribution for these precision parameters with equal emphasis on the spatial and non-spatial component has been proposed (158); however, the value relies on the average number of neighbours, and thus is not generalizable across data (and thus studies) with differing spatial structure. As well, the contribution of each of the two random effects cannot be disentangled (157). A proposed solution is the use of a single precision parameter to capture the variance of the two components, with a mixing parameter estimating the influence of the spatial and non-spatial latent effects (159). In addition to re-parametrizations of the original convolution prior, Riebler *et al.* recently provided an approach to enhance interpretability of the common precision parameter as the marginal of the two random effects through the introduction of scaled spatial structure (160).

3.8 Retail Gravity Analysis

Physical location plays a central role in the success of retail business, as the store's market potential (catchment population) is largely determined by the geographic extent of customers willing to visit the store of interest (161). To generate a map of catchment populations, the study region is partitioned into small and mutually exclusive areas, where each area has an empirically or theoretically derived visit probability of residents corresponding to a set of stores in the surrounding the area. Therefore, each store is thought to have a map of market share: a discrete surface of area-specific store-visit probability reflecting the spatially varying proportion of customers of that store.

The origin (area)-destination (store) pairwise probability is commonly generated by the Huff gravity model, a discrete store choice model parametrized by two influential factors of store visit (162). The law of retail gravitation holds that the probability a resident in an area will use a given store is proportional to the store's business size and inversely proportional to the travel cost to the store, normally defined as travel time or distance (163). The Huff model is widely shown to predict area-level market share accurately in practice (164).

3.9 Ethics Approval

The research protocol in Manuscript 1-3 were reviewed by the Institutional Review Board of the Faculty of Medicine, McGill University and received an expedited approval (IRB Study Number: A07-E45-16B).

Chapter 4: Susceptibility to Price Discounting of Soda by Neighborhood Educational Status: An Ecological Analysis of Disparities in Soda Consumption Using Point-of-Purchase Transaction Data in Montreal, Canada (Manuscript 1)

4.1 Preamble

As described above, there is a dearth of evidence about the role of price discounting in creating nutrition disparity, even though this marketing tactic is acknowledged as one of the most powerful tools to temporarily increase sales. Because discounting is intrinsically time-varying, data collection must be repeated in synchronization with retailers' pricing schedule (typically weekly) on many food items across stores. As the value of weekly pricing and sales are automatically recorded at the time of purchasing by barcode scanning, longitudinal analysis of POS data sheds insights into the potential influence of discounting on community nutrition disparity.

In this manuscript, I investigated the weekly association of soda sales (outcome) and price discounting (exposure) across area-level education in store catchment areas (effect measure modifiers). The manuscript was one of the first studies to document the presence of community nutrition disparity created by price discounting. I used area-level education as the indicator of neighborhood social disadvantage, since accumulating empirical evidence suggests education is a strong determinant of dietary behaviors (more so than income) (22,165).

This manuscript has been accepted and published in the peer-reviewed *International Journal of Epidemiology*

4.2 Manuscript Title Page

Susceptibility to Price Discounting of Soda by Neighborhood Educational Status: An Ecological Analysis of Disparities in Soda Consumption Using Point-of-Purchase Transaction Data in Montreal, Canada.

Hiroshi Mamiya, Erica E.M. Moodie, Yu Ma, David L. Buckeridge

Corresponding to Hiroshi Mamiya, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1140 Ave Pine, Montreal, Quebec, Canada, H3A 1A3 (e-mail: hiroshi.mamiya@mail.mcgill.ca)

Author affiliations: Surveillance Lab, McGill Clinical and Health Informatics, McGill University, Montreal, Quebec, Canada (Hiroshi Mamiya and David L. Buckeridge); Department of Epidemiology Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada (Hiroshi Mamiya, Erica EM. Moodie, David L. Buckeridge); Desautels Faculty of Management, McGill University, Montreal, Quebec, Canada (Yu Ma).

Keywords: Price discounting, Food marketing, Obesity and overweight, Soda, Public health nutrition, Transaction data

4.3 Abstract

Introduction: Price discounting is a marketing tactic used frequently by food industries and retailers, but the extent to which education modifies the effect of discounting on the purchasing of unhealthy foods has received little attention. We investigated whether there was a differential association of price discounting of soda with store-level soda purchasing records between 2008 and 2013 by store-neighborhood education in Montreal, Canada

Methods: Using data on grocery purchase transactions from a sample of supermarkets, pharmacies, supercenters and convenience stores, we performed an ecological time-series analysis, modelling weekly store-level sales of soda as a function of store-level price discounting, store- and neighborhood-level confounders, and an interaction term between discounting and categorical education in the neighborhood of each store.

Results: Analysis by store type (n=18 743, 12 437, 3 965, and 49 533 store-weeks for superstores, pharmacies, supercenters, and convenience stores, respectively) revealed that the effect measure modification of discounting by neighborhood education on soda purchasing was lower in stores in the more educated neighborhoods, most notably in pharmacies (-0.020 [95% CI: -0.028, -0.012] and -0.038 [95%CI: -0.051, -0.025] as an association between log soda sales and percent discounting in proportion, for middle and high education category, respectively). Weaker effect modification was observed in convenience stores. There was no evidence of effect modification in supercenters or superstores.

Conclusions: Price discounting is an important environmental risk factor for soda purchasing and can widen education inequalities in excess sugar intake across levels of education. Interventions to regulate price discounting warrant further investigation as a

public health strategy to improve population nutrition, particularly in lower-education neighborhoods.

4.4 Introduction

Unhealthy eating is the leading preventable cause of death and disability (166). Obesity is one of the main manifestations of excess caloric intake and a major intermediate risk factor for cardiovascular diseases, type 2 diabetes, and many cancers (167). Pervasive marketing of unhealthy foods (i.e. energy-dense, nutrition-poor products) plays an important role in driving the current obesity epidemic (6,16). In particular, carbonated soft drinks (i.e. soda) have the largest marketing expenditure (168) and are the most important source of added sugar intake for adolescents and adults among sugar-sweetened food products in the United States (169).

To date, health research has focused almost exclusively on a single type of food marketing: advertising through communication media, primarily in the form of television commercials targeting children (6,99). Although rarely discussed, marketing activities occurring within stores, such as product display in a prominent location (e.g. end of aisle), in-store flyers, couponing, and price and quantity discounting, are a potentially important marketing channel (16,85,99,170) that accounts for the majority of marketing expenditures by food industries and retailers (171).

Price promotion, or temporary price discounting below regular price (hereafter called <u>price discounting</u>), appears to be the most influential form of marketing (16,172), and unhealthy foods are frequently promoted in this manner (15). Price discounting can induce stockpiling and accelerated consumption of storable and convenience products (16), notably sugar-sweetened beverages (173). Price discounting may also be a critical and overlooked factor explaining socio-economic inequality in dietary patterns if susceptibility to promotion is associated with lower Socio-Economic Status (SES). Moreover, discounting can compromise the effectiveness of taxation targeting sugar-sweetened beverages (102). Thus, given the taxation policies in several countries and jurisdictions, it is important to evaluate the association of discounting with unhealthy food selection, with particular attention to the impact of price discounting on socio-economically disadvantaged populations.(174,175).

The primary challenge in measuring the association between price discounting and purchasing at a population scale is the highly time-varying nature of promotion and purchasing across many food items. Marketing researchers have addressed this challenge by using longitudinal data collected at the retail point-of-purchase (i.e. store level data). These grocery transaction data are collected continuously and automatically from geographically representative samples of retail food stores by marketing companies such as the Nielsen Corporation (176). Although rarely used for population health research, these data provide details of purchased items, including weekly discounting status as well as store locations from which neighborhood SES can be measured, thus offering a potentially rich source of data for research.

Using grocery transaction data, we aimed to estimate the differential association (effect measure modification) of price discounting with unhealthy food purchasing at the store level, as modified by store-neighborhood education. We focused on education rather than other socio-economic characteristics such as income and occupational class, as previous research has shown education to be the main driver of food selection (165,177) and education has demonstrated a consistent association with nutrition-related chronic illness and dietary patterns in studies in Canada and the U.S. (73,178–180). We focused on soda (excluding diet products), since sodas are among the most intensely price-promoted products (15) and their intake has a strong empirical association with weight gain (181).

4.5 Methods

4.5.1 Study Design

We performed an ecological time-series study using store-week as the unit of analysis. Our study area was the Census Metropolitan Area of Montreal (CMA), Quebec, Canada, which includes urban and suburban areas with a population of 3 824 211 inhabitants as of 2011 (182). Data were available on transactions between January 2008 and December 2013, covering a total of six years, or 311 weeks.

4.5.2 Data Source

To create a representative sample of stores in the Montreal CMA, the data provider (Nielsen Corporation) selected chain retail stores by stratified random sampling, with the strata defined by store type (e.g. pharmacy and supermarkets) and size. Some stores were resampled periodically to ensure the sample reflected the changing store mixture. The store location was coded at the level of Forward Sortation Area (FSA), which is a geographical unit defined by the first 3 characters of the 6-character Canadian postal code and utilized in previous Canadian studies (141,142). On average, a single FSA contained 8000 households in 2011 (143).

The sampling frame included four store types: supermarkets (large chain outlets), pharmacies, supercenters (e.g. Wal-Mart), and convenience stores including gas stations, as defined by the Standard Industry Classification codes (183). The sampling frame excluded warehouses such as Costco, and independent (non-chain) stores. Based on comprehensive business establishment data enumerating all retail business in Canada, the market share of our sampling frame among all chain and independent stores selling food items was 65.2% of the CMA.

4.5.3 Transaction Data

The point-of-purchase transaction data provide a weekly, store-level description of products purchased. Retailers typically schedule pricing and other marketing activities for an entire week, hence the data are collected weekly. Variables include product name, category (e.g. regular carbonated soft drinks, milk, fruit juice), unique product identifier, quantity sold, store and chain identification where the purchase took place, store location as FSA, store type and selling price in that week, from which we calculated price discounting as described below.

Additionally, the data record non-price promotions collected by field auditors at supermarkets, pharmacies, and supercenters (but not convenience stores), including weekly display promotion at prominent locations and in-store flyers. We extracted transaction records for products coded as “Carbonated Soft Drinks”, from which diet soda items were removed based on terms suggestive of being sugar-free.

4.5.4 Outcome, Exposure, and Effect Measure Modifier

The outcome was the sum of sales of all soda items at store i on week j, standardized using the Food and Drug Administration standard serving size of 240ml, which was log-transformed due to strong skewness of sales. Because our outcome of interest was overall sales of soda rather than the sales of individual soda items, we followed a conventional modelling approach (184), using store-week level aggregated sales rather than sales of individual products. Item-level analysis accounting for competition of approximately 250 soda items in a typical supermarket leads to many covariates (i.e. approximately 250 for price and another 250 for discounting) that are not fixed across stores. Thus, item-level analysis for transaction data is not feasible with currently available methods (185).

The exposure variable was calculated as a weighted average of discounting of soda items in store *i* on week *j*. The transaction data do not directly measure *regular* (reference, non-discounted) price, instead they provide a record of the price at which items were sold in each store-week. Therefore, we first calculated the regular price per serving in each store-week as the maximum observed price of each item over a 3-month trailing window in accordance with previous literature (170,186). Price discounting of each soda item was then defined as the percent price reduction of the observed price per serving in relation to the regular price per serving in the same store-week. The discounting for each soda item was in turn averaged at the store-week level with weights representing the annual market share (proportion of sales volume) of each item within the soda category (187). A detailed description of this procedure for calculating regular price and discounting is provided in Appendix 1 in Supplementary Material.

Our primary research aim was to determine whether neighborhood education modifies the store-level association between price discounting and sales of soda. We evaluated this aim using a statistical interaction term between price discounting and neighborhood education in regression models. A categorical variable representing neighborhood education was calculated as the proportion of individuals in the FSA having at least a post-secondary certificate or diploma using the 2011 Canadian National Household Survey (139), and values were categorized into three equal groups (tertiles).

4.5.5 Statistical Analysis

We stratified the analysis, as the effect of discounting and its effect modification are expected to differ by store types (188). We used a linear mixed model to evaluate the extent to which store-neighborhood SES modified the effect of price discounting, using a random intercept to account for store-level correlation with sales. We fit the autoregressive correlation structure of order 1 (AR1) to the within-store covariance to account for temporal autocorrelation of residuals.

4.5.6 Confounders

The potential confounders of sales and price discounting that we considered were: regular price per serving, temporal terms (month and year indicator variables), binary indicator variables corresponding to weeks containing a Canadian statutory holiday, categorical indicators for chain identification code, and FSA-level SES and demographic indicators provided in Table 4-1, which are population density per square kilometer, median family income, and the percent of immigrants, families with children, and single parent families. Selection of the covariates was guided by model fit as measured by the Akaike's Information Criteria (AIC).

We performed several sensitivity analyses, considering (i) a stricter definition of price discounting, requiring a discount of at least 5% and (ii) inclusion of product-specific display and flyer promotion indicators, again averaged at the store-week level using the market share of each product as a weight. These promotion variables were available for all store types except convenience stores. We also fit models with area-level education measures calculated from various definitions of store-catchment (Appendix 2, Supplementary Material) and investigated confounding due to across-category competition (e.g. price and discounting of fruits drinks affecting the association between discounting of soda and sales of soda through substitution) (Appendix 3, Supplementary Material).

4.6 Results

In the 6-year study period, there were 18 743, 12 437, 3 965, and 49 533 store-weeks recorded from the 91 supermarkets, 57 pharmacies, 42 supercenters, and 200 convenience stores, respectively. Neighborhood socio-demographic and economic attributes of the sampled stores by store type were similar (Table 4-1). Intensity (percent) of discounting of soda was highest in pharmacies (median percent price reduction: 16.4%, Interquartile Range [IQR] 10.4-22.6) and lowest in convenience stores (5.8%, IQR 3.4-8.6). In general, percent discounting was uniform across neighborhoods within store types (Table 4-2). That is, the data do not suggest that retailers were targeting lower education neighborhoods for discounting.

Figure 4-1 A)–D) show the store-level predicted value of soda sales in response to increasing discounting by the three levels of education. Across all levels of education and types of store, discounting was positively associated with soda sales. The interaction between percent soda discounting (in proportion) and neighborhood education on soda sales (in log serving quantity) was statistically significant for pharmacy and convenience stores (Table 4-3 and Figure 4-2). In general, adjustment of the models with the area-level SES attributes other than income resulted in a lower model fit; thus, only income was retained in all models as the area-level SES confounder.

The modification of the effect of price discounting by education was most prominent in pharmacies (Figure 4-1 b and 4-2), where the average log sales associated with discounting progressively increased as the education tertile decreased. In contrast, the magnitude of effect modification in convenience stores was much smaller (Table 4-3 and Figure 4-2), as seen in the broadly similar patterns of sales response to discounting across education levels (Figure 4-1 d). The confidence interval of the interaction terms in supercenters was widest, reflecting the smallest number of observation (Table 4-3 and Figure 4-2).

Sensitivity analyses using a stricter definition of price discounting (a 5% threshold) or display and in-store flyers included in the model produced nearly identical results (results not shown). The magnitude of the parameter estimates for the interaction terms using expanded store-catchment areas was similar to those from the main analysis, except for supercenters where the parameter estimate tended towards the null value as the catchment area increased in size (Supplementary Figure 4-2). The models including aggregated regular price and discounting of other beverage categories (Supplementary Figure 4-3) gave results consistent with the original estimates, although the effect modification in pharmacies was slightly attenuated.

4.7 Discussion

We used store-level transaction data from a sample of retail food outlets to measure the weekly, time-varying nature of price discounting and to examine whether neighborhood education modified the association between discounting and purchasing of soda. Discounting in pharmacies, and to a lesser degree in convenience stores, was associated with greater increases in purchasing in areas with the lowest educational attainment as compared to areas with higher education. This study is one of the first to explore how education modifies the association of price discounting with purchasing. Despite being one of the most effective food marketing strategies, discounting has received little attention as an environmental risk factor for unhealthy food purchasing and a potential driver of disparities in nutrition that can cause chronic diseases (15).

The descriptive analysis revealed a nearly uniform intensity of price discounting by area-level educational attainment *within* each store type. This finding is consistent with a previous report assessing the pattern of price discounting by area SES using a nationally representative sample of stores in the United States (15). *Across* store types, however, the greater discounting for soda in pharmacies and supercenters as compared to supermarkets calls attention to the emerging role of food retailers *other than* supermarkets and convenience stores as important sources of potentially unhealthy food purchases.

The magnitude of the interaction between price discounting and education varied substantially by store type, with the effect measure modification the greatest for pharmacies. It is possible that higher intensity of price discounting in pharmacies relative to other store types attracts "deal-prone" individuals (188) who may be more prevalent in low-education areas. Despite having a lower market share than supermarkets, pharmacies are a critical component of the retail store mix due to their broad geographic accessibility, especially in areas with limited physical access to supermarkets. A similar pattern of effect modification was observed in convenience stores, although to a smaller extent. In contrast, high and middle education neighborhoods had a greater response to price discounting in supermarkets than the lowest (reference) education category, although the confidence intervals crossed the null value and thus the effect modification was inconclusive.

The characteristics of population served by each store type are not well known; however, a growing proportion of consumers shop for groceries at multiple types of stores, regardless of SES (189,190). Therefore, although supermarkets supply most groceries today, our study suggests the importance of including non-traditional retail settings such as pharmacies for investigation of in-store retail food environment and interventions to improve population nutrition. A previous study demonstrating an association between body mass index and sales promotions of energy-dense food products used cross-sectional measures of promotion manually collected from food establishments in a highly confined geographical area (i.e. a "food desert" neighborhood) (85). A previous cross-sectional study based on a nationally representative survey of households in the United Kingdom focused on large supermarkets with results similar to our superstore subsample (17). However, the study failed to address highly variable sensitivity to discounting by store types (188).

Manual and longitudinal in-store observation of many food items from multiple store types is prohibitively expensive at a large geographic scale. Because objective data collection was achieved using scanner technology at the store-level in our study, the purchase and promotional status are free of measurement error arising from manual store audits or reporting from participants. Our study thus makes a valuable methodological contribution to the measurement of in-store marketing from a public health perspective.

4.7.1 Limitations

Although our sensitivity analysis demonstrated the robustness of the results to different definitions of store catchment area, our measure of education may be subject to misclassification for supercenters, which are expected to have the largest catchment areas. The findings of store-level association cannot be interpreted at the individual shopper level. However, while the use of store scan data to capture weekly time-varying measures from a representative sample of stores requires an ecological analysis, we believe that this limitation is outweighed by the benefits of these data. Third, our sample does not include independent (i.e. non-chain) stores, which may predominate in some neighborhoods despite having a smaller share of the food market overall (35%). However, chain stores have a substantially higher intensity of discounting relative to independent stores (15); thus, we included the most relevant store type with which to study the impact of price discounting. Finally, because the transaction data were generated by a sample of stores, we were unable to investigate potential substitution of soda sales among nearby stores due to price competition, although across-store competition is typically limited to high-cost products (e.g. coffee and beer), and the proportion of such “cherry-picking” customers is typically small, ranging between 10-15% of all shoppers (191,192).

4.7.2 Future Work

Although we focused on a single food category, our methodology can be extended to all (healthy and unhealthy) food categories. Doing so will allow exploration of whether discounting causes substitution to healthier products or categories (99) as well as assessment of differential price sensitivity between healthy and unhealthy food categories (17).

4.7.3 Implications

Our findings suggest that interventions to regulate price discounting warrant further investigation as a public health strategy to reduce population-wide consumption of soda, and to potentially narrow socio-economic gaps in excess sugar intake. We stress, however, the complementary nature of potential interventions aimed at discounting to other nutrition policies, including subsidization of healthy foods and taxation of unhealthy food, since sustained behavioral change requires multi-component interventions (193,194). Given the growing trend of grocery shopping at multiple store types (195) and the differential susceptibility to price discounting by store type as demonstrated in this study, planning and monitoring the effectiveness of intervention should occur at a population scale for all store types.

Although evidence from a food-environment study such as ours is not necessarily portable from one population to another, our findings may inform future hypothesis for subsequent investigation of price discounting. Furthermore, we demonstrated an analytical approach that is applicable to point-of sales transaction data, which are widely available for research and public health practice.

Finally, we note that pharmacies can be an important point of interest for health promotion and chronic disease management due to their ubiquity and the motivation of pharmacists to adopt public health initiatives beyond medication management (196). However, our study indicates that pharmacies can expose shoppers to an unhealthy selection of food items. Health agencies should aid pharmacies in taking leadership roles in health promotion by creating environments for healthy eating.

Table 4-1. Distribution of Soda Sales, Discounting, Price, and Socio-Demographic and Economic Attributes by Sampled Store Type in Greater Montreal

	Supermarket (n=91, N=18 743)					Pharmacy (n=57, N=12 437)				
	Min	**Mean**	**Median**	**IQR**	**Max**	**Min**	**Mean**	**Median**	**IQR**	**Max**
Store-level characteristic†										
Log sales of soda in servings	7.3	10.1	10	(9.5-10.7)	12.6	0.4	6.8	7	(6-7.9)	10.5
Percent price discounting	0	11.7	11.2	(8.1-14.7)	34.8	0	16.7	16.4	(10.4-22.6)	65.4
Regular price±	21.4	36.1	35.8	(31.2-40.9)	58.6	17.4	48.6	47.4	(41-55.2)	95.5
Net price ξ	18	32.5	32.1	(28.1-36.6)	55.6	10.7	42.2	40.6	(34.2-49)	95.5
Percent of post-secondary diploma or certificate or higher	43.5	67.7	66.7	(60.5-76.9)	89.4	49.1	67.2	66.6	(58.9-75.6)	89.4
Area-level characteristic§										
Median family income in 10,000 Canadian dollars	3.8	7.3	7	(5.6-8.5)	14.2	4.2	7.2	7.2	(5.9-8.2)	14.2
Percent immigrants	2.7	21.3	16.2	(6.2-34.4)	60.6	1.5	21.7	22.6	(7.8-35)	54.3
Percent family with child	24.5	43.9	45.5	(38.7-49.8)	66.7	26.3	42.5	42.9	(37-46.6)	60.1
Percent single parent family	9.8	17.1	16.4	(13.9-20.4)	27.4	6.5	17.4	16	(14.1-20.4)	27.4
Population density per square kilometre	146.3	3718.5	2603.8	(969-5507.7)	15211.3	148.2	3588.7	2368.4	(1360.7-6065.1)	12195.5

Abbreviation: IQR, Interquartile range: n, Number of sampled store: N, Number of store-weeks
Min: Minimum, Max: Maximum
† Store-level characteristics are time-varying and generated from store-week observations
§ Area-level characteristics are time-invariant and calculated from areas where stores are located
± None-discounted price (in cents) per serving of soda, defined as the maximum price of each item over a 3-month trailing window in Canadian Cents
ξ Price after discounting per serving of soda in Canadian Cents

Table 4-1 continued

	Supercenter (n=42, N=3 965)					Convenience store (n=200, N=49 533)				
	Min	**Mean**	**Median**	**IQR**	**Max**	**Min**	**Mean**	**Median**	**IQR**	**Max**
Store-level characteristic†										
Log sales of soda in servings	2.2	8.9	9.3	(7.8-9.9)	11.8	1.8	6.7	6.8	(6.3-7.3)	8.8
Percent price discounting	0	12.4	11.6	(7.8-16.5)	35.3	0	6.2	5.8	(3.4-8.6)	27.1
Regular price±	15.5	36.1	32.8	(29.6-42.5)	66	40.6	68.2	67.3	(62.4-73.5)	164.4
Net price$^{\xi}$	10.8	32.6	30.1	(27.1-37.7)	63.9	36.3	64.1	63.8	(58.8-69.5)	124.2
Percent of post-secondary diploma or certificate or higher	53.2	68.3	66.6	(63.5-76.9)	84.8	48.2	67.8	67.5	(61.1-74.6)	92
Area-level characteristic§										
Median family income in 10,000 Canadian dollars	4.2	7.6	7.9	(6.2-8.9)	10.3	4.2	7.1	7	(5.4-8.3)	12.9
Percent immigrants	2.7	18.6	18.4	(8.6-30)	54.3	1.4	23.6	21.9	(9.4-35.9)	63.8
Percent family with child	30.3	44.4	42.9	(34.1-52.7)	60.1	17.5	43.2	45.2	(36.5-49.5)	66.7
Percent single parent family	11.6	16	14.1	(11.8-19.8)	23.7	7.9	17.6	16.7	(14.3-20.4)	28.8
Population density per square kilometre	326.1	2029.2	1737.3	(1338.6-2764.6)	12040.1	37.8	3459.2	2882.5	(1051.8-5050.2)	10510.1

Abbreviation: IQR, Interquartile range: n, Number of sampled store: N, Number of store-weeks
Min: Minimum, Max: Maximum
† Store-level characteristics are time-varying and generated from store-week observations
§ Area-level characteristics are time-invariant and calculated from areas where stores are located
± None-discounted price (in cents) per serving of soda, defined as the maximum price of each item over a 3-month trailing window in Canadian Cents
$^{\xi}$ Price after discounting per serving of soda in Canadian Cents

Table 4-2. Distribution of Percent Price Discounting of Soda by Sampled Store Type and Neighborhood Education

Price discounting (percent)		All store types (n=84 678)		Supermarket (n=18 743)		Pharmacy (n=12 437)		Supercenter (n=3965)		Convenience store (n=49 533)	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR
Education tertile[a]	Low	8.2	(4.7 - 12.8)	11.4	(8.0 - 14.2)	17.2	(10.6 – 23.0)	16.2	(11.9 - 20.4)	6.0	(3.6 - 8.8)
	Middle	7.7	(4.5 - 12.1)	11.8	(8.7 - 15.1)	17.0	(10.8 - 23.7)	10.6	(7.2 - 15.4)	5.7	(3.3 - 8.4)
	High	7.8	(4.5 - 12.1)	10.8	(7.6 - 14.5)	15.1	(9.8 - 20.9)	10.2	(7.6 - 15.3)	5.7	(3.3 - 8.5)
All areas[b]		7.9	(4.6 - 12.3)	11.2	(8.1 - 14.7)	16.4	(10.4 - 22.6)	11.6	(7.8 - 16.5)	5.8	(3.4 - 8.6)

Abbreviation: IQR, Interquartile range
n, number of store-weeks
[a] Tertile area-level proportion of resident holding at least post-secondary certificate or diploma
[b] Value from combined education levels

Table 4-3. Regression Coefficient of Percent Price Discounting, Education, Their Interaction, and Intercept by Store Type

	Supermarket (n=18 743) Estimate (95% CI)	**Pharmacy (n=18 743) Estimate (95% CI)**	**Supercenter (n=18 743) Estimate (95% CI)**	**Convenience store (n=18 743) Estimate (95% CI)**
Intercept	9.584 (9.342, 9.827)	6.428 (6.065, 6.792)	7.040 (6.719, 7.361)	6.482 (6.237, 6.727)
Percent discounting [a]	0.038 (0.036, 0.040)	0.136 (0.130, 0.142)	0.051 (0.042, 0.059)	0.028 (0.025, 0.030)
Education (middle) [b]	0.023 (-0.215, 0.260)	0.144 (-0.259, 0.548)	0.041 (-0.382, 0.464)	-0.220 (-0.433, -0.007)
Education (high) [b]	-0.113 (-0.348, 0.123)	0.580 (0.150, 1.010)	-0.097 (-0.628, 0.434)	-0.125 (-0.368, 0.118)
Percent discounting X Education (middle)	0.001 (-0.002, 0.004)	-0.020 (-0.028, -0.012)	-0.005 (-0.019, 0.008)	-0.005 (-0.008, -0.001)
Percent discounting X Education (high)	0.001 (-0.002, 0.003)	-0.038 (-0.051, -0.025)	-0.010 (-0.033, 0.013)	-0.007 (-0.010, -0.003)

Abbreviation: 95% CI, 95% Confidence Interval; X, Two-way interaction (product term)

[a] Average store-level discounting of soda in percent. The value is interpreted as a change in natural log soda sales in serving associated with one percent discounting of soda in a store located at lowest tertile of education.
[b] Categorical variable for area-level proportion of resident holding at least post-secondary certificate or diploma.

Note: Area-level SES attributes other than income (i.e. population density per square kilometer and the percent of immigrants, family with children, and single parent family) were not retained in the model.
Coefficient of banner (chain), regular price, income, month, year, and holidays are not displayed.

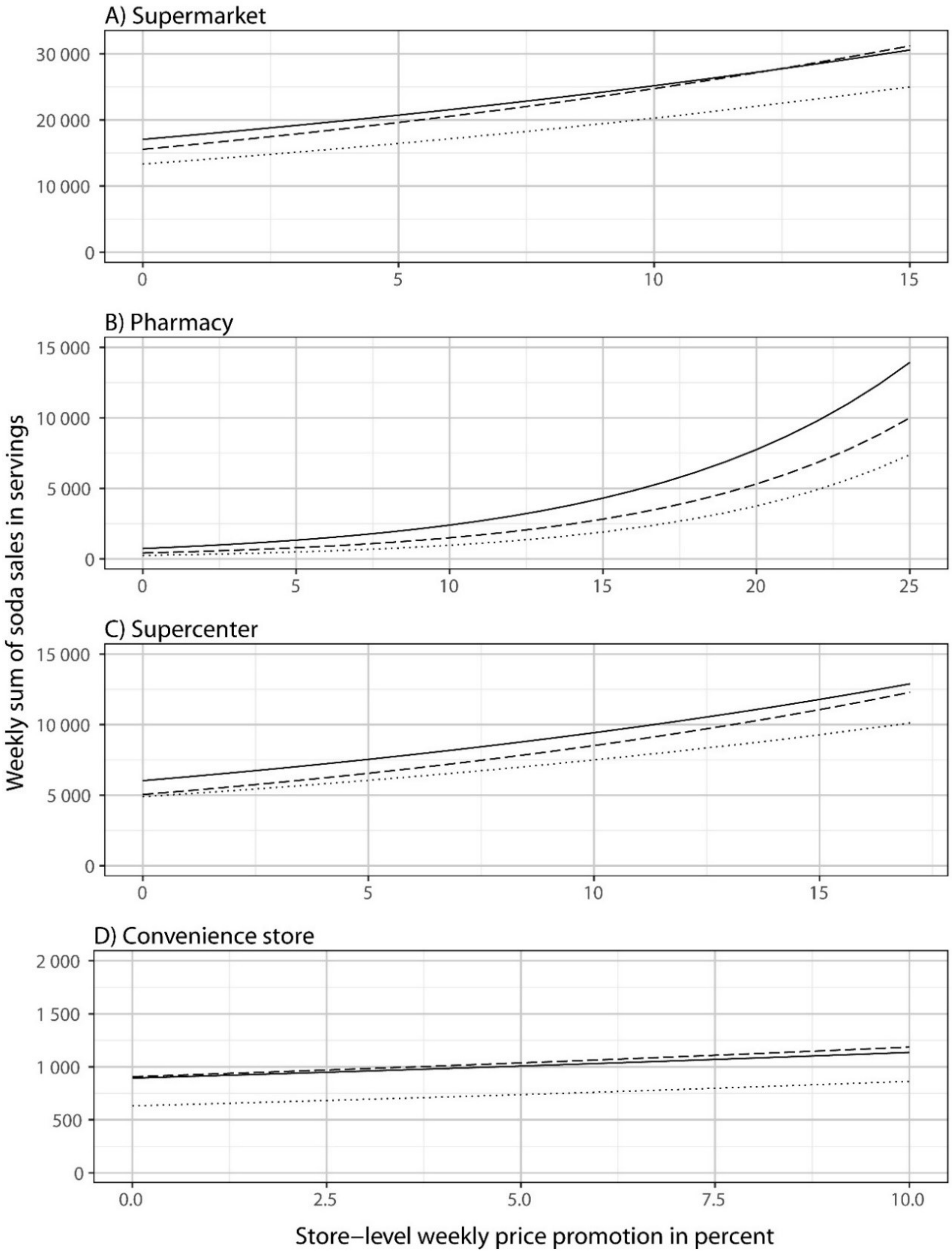


Figure 4-1 A)-D). Predicted Soda Sales (in non-log scale) as a Function of Neighbourhood Education and Percent Price Discounting in the Subsample of A) Supermarket, B) Pharmacy, C) Supercentre, D) Convenience Store

Tertile proportion of area-level education (post-secondary certificate or diploma); dotted line: high education; dashed line, middle education; solid line, low education.

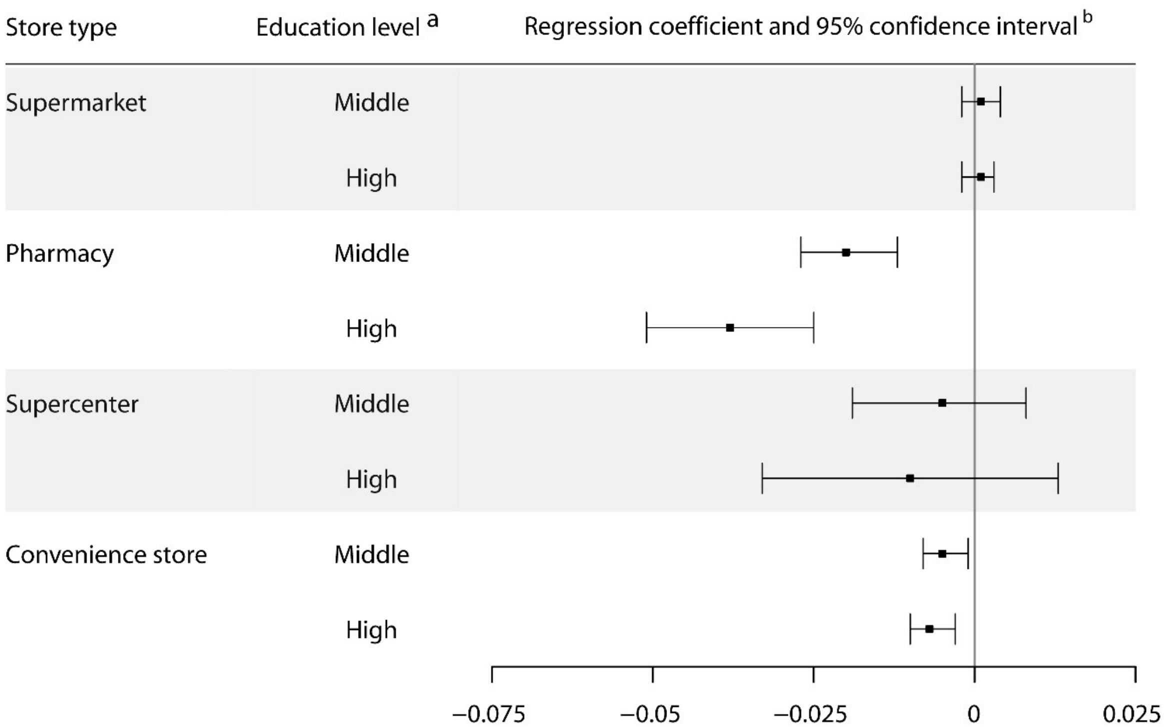


Figure 4-2. Plot of Coefficients for the Interaction of Percent Price Discounting and Neighbourhood Education

[a] interaction of percent discounting with categorical dummy variable for neighbourhood education, using lowest education level as the reference;
[b] association with natural log soda sales (in serving)

4.8 Supplementary Materials for Manuscript 1

Appendix 1. Detailed Description of Calculation of Category-Level Discounting

We used a conventional marketing approach to calculate the (soda) category-level discounting for each store-week from store-level scanner data (1).

The terms used in the analysis are defined as:

1) *Item* refers to a unique soda product as identified by Universal Product Code in transaction data. A medium-sized supermarket typically carries approximately 250 distinct soda items.

2) *Net price* indicates the price of a soda item after discounting, standardized to a single serving size (240ml).

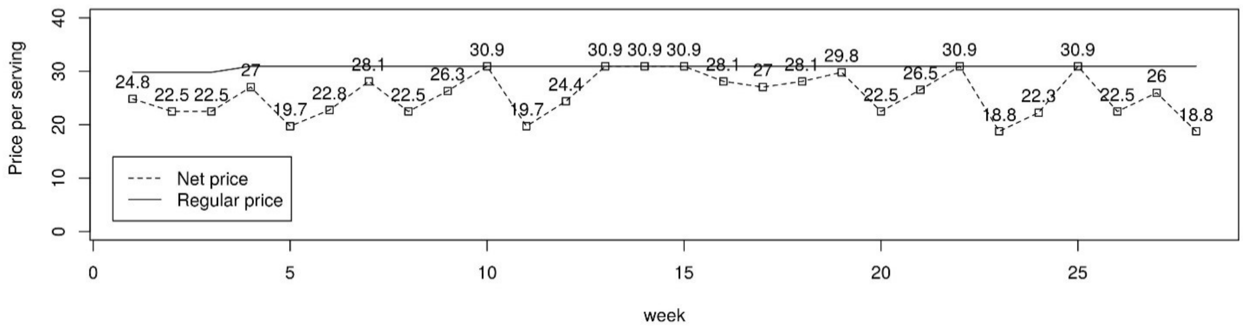
3) *Regular price* is the baseline (reference) price of a single serving without any discounting, and

4) *Item discounting* indicates the difference between the net and regular price, thus the extent or depth of temporary price reduction per serving. Note that net and regular price are in Canadian Cents.

Because the data provide only the weekly net price after discounting for each item but not the regular price nor discounting directly, we calculated the regular price for each item, and then we computed discounting as the difference between the calculated regular price and the observed net price. We then computed the category-level discounting as the weighted mean of item discounting across items in the same store-week, with the weights representing the market share of each item.

Let i represent item of soda sold in store s at week j. We denote the net price of item i in store s

at week j as P_{ijs}. We examined the price history of each item. The regular price of soda item R_{ijs} is identified as the highest price in a 3-month window, e.g., a moving maximum net price of item i in store s in a 3 months widow (1). Supplementary Figure 4-1 below illustrates a series of weekly net price of one of popular soda item (12 units X 355ml of a major soda brand) and its regular price.



Supplementary Figure 4-1. Weekly Net Price and Regular Price of a Soda Item in Cents

Item discounting, D_{ijs} , is defined as a percent price decrease from regular price as shown below;

D_{ijs}= $(R_{ijs} - P_{ijs})$ / $R_{ijs} \times 100$

Supplementary Figure 4-1 indicates that there is no item discounting (zero percent item discounting) in weeks 10, 13-15, 23, and 25, so the net price is the same as the regular price for these weeks. In other weeks, there is discounting since the net price is smaller than the regular price. In week 11, for example, the discounting is computed as: (30.9 – 19.7) / 30.9 x 100 = 36.25 percent.

Finally, we aggregated the individual item discounting into an overall measure for the discounting in the soda category. Suppose there are 300 soda items sold in store s in week j. Our main exposure of interest, the store-week category discounting of soda, X_{js},is defined as

$$X_{js} = \sum_{i=1}^{300} D_{ijs} W_{is}$$

,

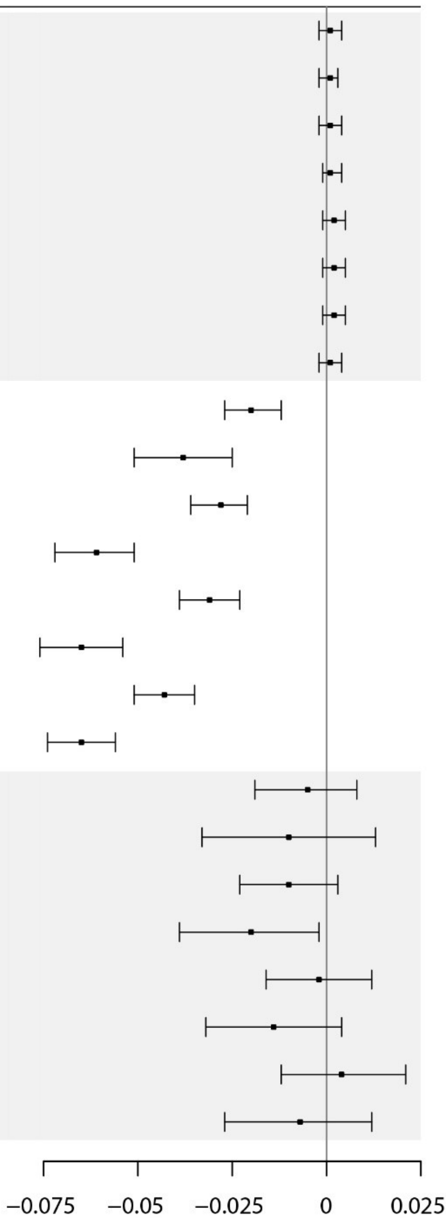
where W_{is} represents a normalized (sum to one across all items in store s) weight calculated from the annual market share of each soda items at store s. Therefore, the value of the weights is large and stable over time for soda items of major brands, such as Pepsi and Coca-Cola.

As described in the main text, we also performed a sensitivity analysis with a stricter definition of discounting, where item discounting (D_{ijs}) was set to zero before calculating X_{js} if the percentage decrease in price was less than 5%. i.e., we only consider a price decrease of more than 5% to be discounting. The results are consistent with the original definition of category discounting above.

Appendix 2. Sensitivity Analysis for Area-Level Education

In the main analysis, store neighborhood education was measured at the level of the Forward Sortation Area (FSA), which is the geocoded spatial location of store in the transaction data. To assess the robustness of our estimates based on various definition of store catchment area, we re-fit the model with area-level education calculated from expanded catchment areas around the original FSA, by adding buffer zones of 0.5 km, 1.0 km, and 2.0 km, thereby creating smoothed measures accounting for education in surrounding FSAs. Because convenience stores are expected to have much smaller catchment area than other store types, we did not apply this sensitivity analysis to these stores. The results are provided in Supplementary Figure 4-2. The magnitude of the parameter estimates for the interaction terms was similar to those from the main analysis, except for supercenters where the parameter estimate shifted towards the null value as the catchment area increased in size.

Store type	Area[a]	Education level	Estimate	95% CI
Supermarket	Original	Middle	0.001	(−0.002 , 0.004)
		High	0.001	(−0.002 , 0.003)
	0.5km	Middle	0.001	(−0.002 , 0.004)
		High	0.001	(−0.001 , 0.004)
	1.0km	Middle	0.002	(−0.001 , 0.005)
		High	0.002	(−0.001 , 0.005)
	2.0km	Middle	0.002	(−0.001 , 0.005)
		High	0.001	(−0.002 , 0.004)
Pharmacy	Original	Middle	−0.020	(−0.027 , −0.012)
		High	−0.038	(−0.051 , −0.025)
	0.5km	Middle	−0.028	(−0.036 , −0.021)
		High	−0.061	(−0.072 , −0.051)
	1.0km	Middle	−0.031	(−0.039 , −0.023)
		High	−0.065	(−0.076 , −0.054)
	2.0km	Middle	−0.043	(−0.051 , −0.035)
		High	−0.065	(−0.074 , −0.056)
Supercenter	Original	Middle	−0.005	(−0.019 , 0.008)
		High	−0.010	(−0.033 , 0.013)
	0.5km	Middle	−0.010	(−0.023 , 0.003)
		High	−0.020	(−0.039 , −0.002)
	1.0km	Middle	−0.002	(−0.016 , 0.012)
		High	−0.014	(−0.032 , 0.004)
	2.0km	Middle	0.004	(−0.012 , 0.021)
		High	−0.007	(−0.027 , 0.012)



Supplementary Figure 4-2. Plot of Coefficients for the Interaction of Percent Price Discounting and Neighborhood Education Defined by Various Size of Store Catchment Area

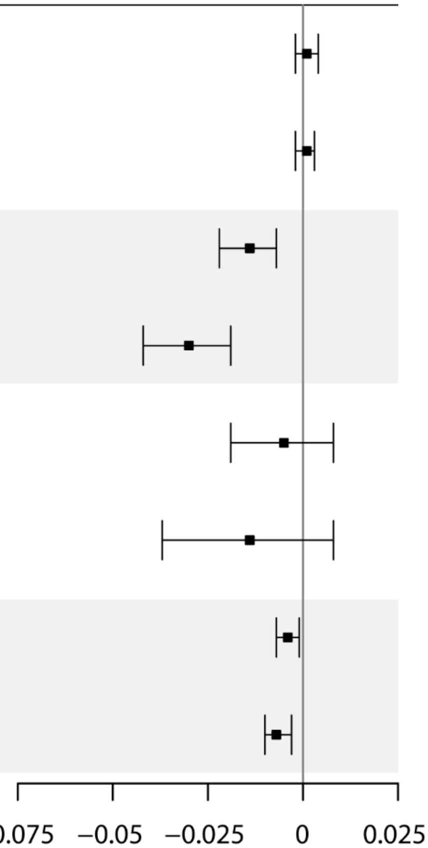
Abbreviation: 95% CI, 95% Confidence Interval

[a] Definition of store catchment area where education measures are captured.
Original; Original definition of neighborhood (Forward Sortation Area)
0.5km; 0.5km buffer zone around the original FSA
1.0km; 1.0km buffer zone around the original FSA
2.0km; 2.0km buffer zone around the original FSA
Note: The analysis was not performed for convenience store, since they are likely to have catchment area much smaller than FSA.

Appendix 3. Sensitivity Analysis Addressing Competition With Other Beverage Categories.

Because association of soda discounting with soda sales may be confounded by price and promotion of other beverage categories such as diet soda, we added covariates representing aggregated category-level regular price and discounting of potentially competing beverage categories, including diet soda, fruits drinks, sports/energy-drinks, frozen fruit drinks, and fruit juices (100% fruit beverage). Since most pharmacies and convenience stores did not sell frozen fruits beverage and fruits juice, these categories were excluded from the model for pharmacy and convenience store. Aggregated category-level regular price and discounting of these beverage categories were computed using the same method used for the soda category in the main analysis. The result suggests minimal confounding by other beverage categories on the interaction of area-level education and discounting (Supplementary Figure 4-3). The only possible exception is for pharmacies, where the magnitude of interaction terms decreased slightly, from (-0.020 [95% CI: -0.028, -0.012] and -0.038 [95%CI: -0.051, -0.025] in the original analysis to -0.014 [95%CI: -0.022, -0.007] and -0.030 [95%CI: -0.042, -0.019] for the middle and high education category, respectively).

Store type	Education level	Estimate	95% CI
Supermarket	Middle	0.001	(−0.002 , 0.004)
	High	0.001	(−0.002 , 0.003)
Pharmacy	Middle	−0.014	(−0.022 , −0.007)
	High	−0.030	(−0.042 , −0.019)
Supercenter	Middle	−0.005	(−0.019 , 0.008)
	High	−0.014	(−0.037 , 0.008)
Convenience store	Middle	−0.004	(−0.007 , −0.001)
	High	−0.007	(−0.01 , −0.003)



Supplementary Figure 4-3. Plot of Coefficients for the Interaction of Percent Price Discounting and Neighborhood Education, conditional on regular price and discounting of beverage categories other than soda

Abbreviation: 95% CI, 95% Confidence Interval

Appendix 4. Analysis Restricted to Pepsi and Coca-Cola Brand.

We first selected items belonging to Coca-Cola and Pepsi brands by identifying items associated with these brands in their item descriptions in the transaction data. Among 851 distinct soda items in the transaction data in the Greater Montreal during the study period, 59 items belonged to Coca-Cola and 85 items belonged to Pepsi. As an example, items belonging to the Coca-Cola brand included a 2.0L bottle, a single 355 mL can, and a pack of 6 355 mL cans. Together, the two brands accounted for the majority of soda sales (54%, 70%, 67%, and 65% in terms of sales volume for supermarket, pharmacy, supercenter, and convenience store, respectively).

In a manner similar to Table 4-1 in the main text, we provide the distribution of pricing and discounting of Coca-Cola and Pepsi branded soda in Supplementary Table 4-1. The regular price and net price of Coca-Cola and Pepsi branded soda are slightly higher than those of the original category including all soda items in supermarket and supercenter. The magnitude of discounting was modestly higher in Coca Cola and Pepsi brands in all store types, except convenience store.

We conducted the same analysis estimating the interaction of price discounting and neighbourhood education for each store type using only Coca-Cola and Pepsi branded soda. The results are consistent with the findings from the main analysis using all soda items, except that the interaction term for the middle education category in convenience stores became inconclusive (95% CI crossing the null) (Supplementary Table 4-2 and Supplementary Figure 4-4). Finally, the main model term for discounting (the association of discounting and purchasing at the lowest education level) increased substantially in supermarket and pharmacy, suggesting a stronger association for the major brands. This finding is consistent with marketing studies that show promotions of larger brands produce greater effects on purchasing than promotions of smaller brands (2,3).

Supplementary Table 4-1. Distribution of Sales, Discounting and Pricing of Coca-Cola and Pepsi Brand

	Supermarket (n=91, N=18 743)					Pharmacy (n=57, N=12 437)				
	Min	**Mean**	**Median**	**IQR**	**Max**	**Min**	**Mean**	**Median**	**IQR**	**Max**
Log sales of soda in servings	4.6	8.7	8.7	(8.1-9.4)	11.7	0.4	5.9	6	(4.8-7.1)	9.9
Percent price discounting	0	12.5	12	(7.2-17.2)	43.3	0	18.6	18.1	(9.6-26.6)	66.1
Regular price ±	23.9	39.8	39.6	(36.6-42.8)	58.7	15.1	45.7	44.8	(39.8-50.3)	93
Net price $^{\xi}$	19	35.7	35.4	(32.8-38.3)	53.9	9.1	38.7	37.5	(32.3-43.6)	93

Abbreviation: IQR, Interquartile range: n, Number of sampled store: N, Number of store-weeks
Min: Minimum, Max: Maximum

$^{\pm}$ None-discounted price (in cents) per serving of soda, defined as the maximum price of each item over a 3-month trailing window in Canadian Cents.

$^{\xi}$ Price after discounting per serving of soda in Canadian Cents.

Supplementary table 4-1 continued

	Supercenter (n=42, N=3 965)					Convenience store (n=200, N=49 533)				
	Min	**Mean**	**Median**	**IQR**	**Max**	**Min**	**Mean**	**Median**	**IQR**	**Max**
Log sales of soda in servings	2.1	7.9	8.1	(6.8-9)	11.3	0.4	5.7	5.8	(5.2-6.3)	8.2
Percent price discounting	0	14.1	13.6	(8.2-19.2)	51.3	0	6.2	5.4	(2.2-9.3)	49.1
Regular price ±	18.2	38	37.1	(34.3-41.1)	64.7	42.2	65.1	65.3	(59.8-70)	144.7
Net price ξ	11.6	33.8	33.4	(30.8-36.9)	62.1	37.3	61.4	61.1	(56.1-66.1)	107.7

Abbreviation: IQR, Interquartile range: n, Number of sampled store: N, Number of store-weeks
Min: Minimum, Max: Maximum

± None-discounted price (in cents) per serving of soda, defined as the maximum price of each item over a 3-month trailing window in Canadian Cents.

ξ Price after discounting per serving of soda in Canadian Cents.

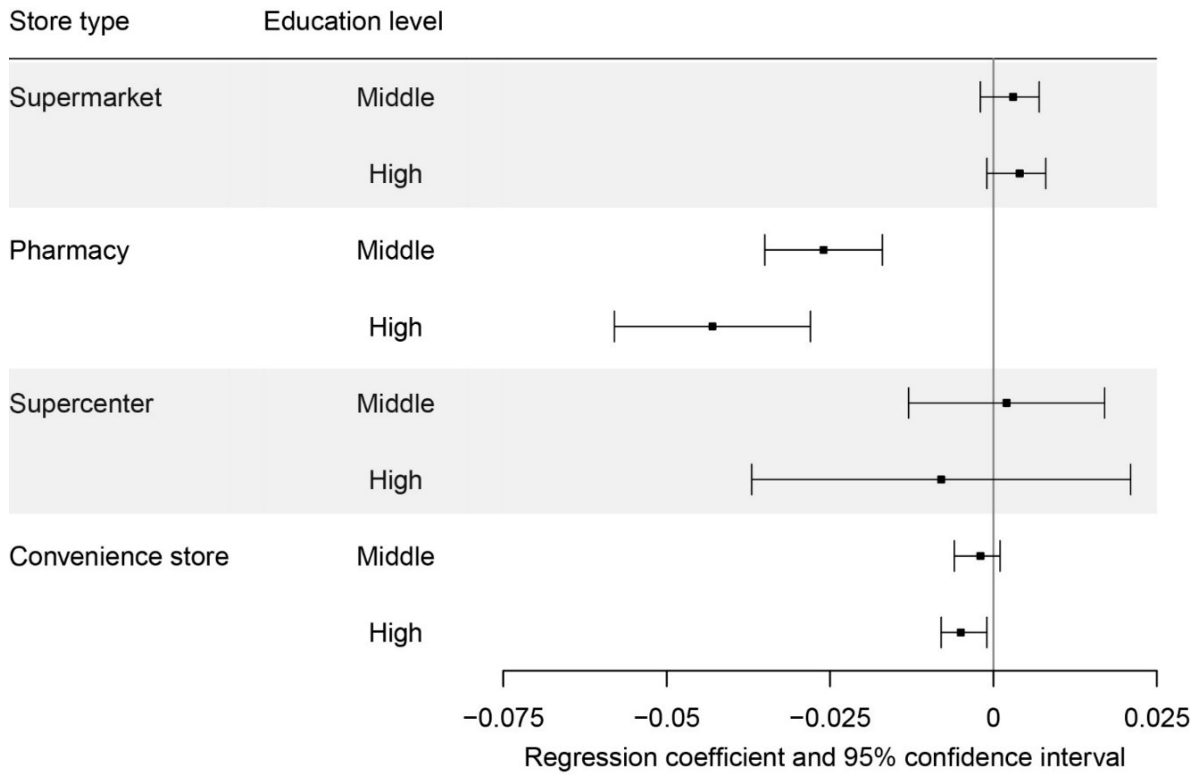
Supplementary Table 4-2. Regression Coefficient of Percent Price Discounting, Education, Their Interaction, and Intercept for Coca Cola and Pepsi Brand

	Supermarket (n=18 743) Estimate (95% CI)	**Pharmacy (n=12 437) Estimate (95% CI)**	**Supercenter (n=3 965) Estimate (95% CI)**	**Convenience store (n=49 533) Estimate (95% CI)**
Intercept	8.675 (8.416 , 8.933)	5.975 (5.621 , 6.328)	6.581 (6.247 , 6.916)	6.166 (5.916 , 6.415)
Percent discounting[a]	0.069 (0.066 , 0.072)	0.157 (0.150 , 0.164)	0.051 (0.041 , 0.061)	0.027 (0.024 , 0.029)
Education (middle)[b]	0.010 (-0.238 , 0.257)	0.121 (-0.270 , 0.513)	0.018 (-0.445 , 0.480)	-0.236 (-0.454 , -0.018)
Education (high)[b]	-0.177 (-0.426 , 0.072)	0.559 (0.140 , 0.978)	-0.164 (-0.729 , 0.401)	-0.170 (-0.418 , 0.078)
Percent discounting X Education (middle)	0.003 (-0.002 , 0.007)	-0.026 (-0.035 , -0.017)	0.002 (-0.013 , 0.017)	-0.002 (-0.006 , 0.001)
Percent discounting X Education (high)	0.004 (-0.001 , 0.008)	-0.043 (-0.058 , -0.028)	-0.008 (-0.037 , 0.021)	-0.005 (-0.008 , -0.001)

Abbreviation: 95% CI, 95% Confidence Interval; X, Two-way interaction (product term)

[a] Average store-level discounting of soda in percent. The value is interpreted as a change in natural log soda sales in serving associated with one percent discounting of soda in a store located at lowest tertile of education.
[b] Categorical variable for area-level proportion of resident holding at least post-secondary certificate or diploma.

Note: Area-level SES attributes other than income (i.e. population density per square kilometer and the percent of immigrants, family with children, and single parent family) were not retained in the model.
Coefficient of banner (chain), regular price, income, month, year, and holidays are not displayed.



Supplementary Figure 4-4. Plot of Coefficients for the Interaction of Percent Price Discounting and Neighborhood Education for Soda Category Limited to Coca-Cola and Pepsi Brand

Abbreviation: 95% CI, 95% Confidence Interval

References for Supplementary Materials

1. Raju JS. The Effect of Price Promotions on Variability in Product Category Sales. Mark Sci [Internet]. 1992 [cited 2016 Apr 3];11(3):207–20. Available from: http://www.jstor.org/stable/183886

2. Blattberg RC, Wisniewski KJ. Price-Induced Patterns of Competition. Mark Sci [Internet]. 1989 Nov [cited 2018 Mar 20];8(4):291–309. Available from: http://pubsonline.informs.org/doi/abs/10.1287/mksc.8.4.291

3. Allenby GM, Rossi PE. Quality Perceptions and Asymmetric Switching Between Brands. Mark Sci [Internet]. 1991 Aug 1 [cited 2018 Mar 20];10(3):185–204. Available from: https://pubsonline.informs.org/doi/abs/10.1287/mksc.10.3.185

Chapter 5: An Area-Level Indicator of Latent Soda Demand: Spatial Statistical Modeling of Grocery Store Transaction Data to Characterize the Nutritional Landscape in Montreal, Canada (Manuscript 2)

5.1 Preamble

Following the temporal analysis in Manuscript 1, I develop a spatial statistical model of POS data to generate a small area indicator of (latent) soda sales at the spatial scale of neighborhood in the CMA of Montreal. The dissemination of this study was timely, given the need for a neighbourhood indicator of diet to allow effective development and tracking of community interventions. Because direct estimation of small-area diets through an increased sample size of nutrition survey is infeasible, model-based estimation using routinely and automatically generated POS data is a viable solution to this measurement gap. This is the first public health research to exploit electronic grocery sales data to learn about community diets, offering a far finer spatial resolution than the currently available local measures.

Since not all small areas contained a store, there was a need to learn information of soda sales from neighboring areas. This challenge led to an analytical approach that accounted for residents' shopping activities in neighboring areas via spatial smoothing in a hierarchical Bayesian model. Following the generation of the indicator, I demonstrated its utility as a predictor for disease mapping of T2D. The indicator provided a better model fit relative to the model containing previously used proxy measure of diets, community food environment measure.

This manuscript has been accepted and published in the peer-reviewed *American Journal of Epidemiology*.

5.2 Manuscript Title Page

An Area-Level Indicator of Latent Soda Demand: Spatial Statistical Modeling of Grocery Store Transaction Data to Characterize the Nutritional Landscape in Montreal, Canada.

Hiroshi Mamiya, Alexandra M. Schmidt, Erica EM. Moodie, Yu Ma, and David L. Buckeridge

Corresponding to Hiroshi Mamiya, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1140 Ave Pine, Montreal, Quebec, Canada, H3A 1A3 (e-mail: hiroshi.mamiya@mail.mcgill.ca)

Author affiliations: Surveillance Lab, McGill Clinical and Health Informatics, McGill University, Montreal, Quebec, Canada (Hiroshi Mamiya and David L. Buckeridge); Department of Epidemiology Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada (Hiroshi Mamiya, Alexandra M. Schmidt, Erica EM. Moodie, David L. Buckeridge); Desautels Faculty of Management, McGill University, Montreal, Quebec, Canada (Yu Ma).

Keywords: Diabetes Mellitus, Disease Mapping, Ecological Analysis, Grocery Transaction Data, Nutrition, Public Health Surveillance

5.3 Abstract

Measurement of neighborhood dietary patterns at high spatial resolution allows public health agencies to identify and monitor communities with an elevated risk of nutrition-related chronic diseases. Currently, data on diet are obtained primarily through nutrition surveys, which produce measurements at low spatial resolutions. The availability of store-level grocery transaction data provides an opportunity to refine the measurement of neighborhood dietary patterns. We used these data to develop an indicator of area-level latent demand of soda in the Census Metropolitan Area of Montreal in 2012 by applying a hierarchical Bayesian spatial model to data on soda sales from 1,097 chain retail food outlets. The utility of the indicator of latent soda demand was evaluated by assessing its association with the neighborhood relative risk of prevalent Type 2 Diabetes. The indicator improved the fit of the disease mapping model (deviance information criterion = 2140 with the indicator and 2148 without) and enables a novel approach to nutrition surveillance.

5.4 Introduction

Unhealthy diet is the most important preventable cause of global death and disability (33). Excess intake of artificially added sugar and fat through energy-dense and nutrition-poor ('junk') food, such as soda (carbonated soft drinks), is a major driver of the global obesity epidemic and nutrition-related chronic diseases, including Type 2 Diabetes Mellitus (T2D) (181). Food choices vary geographically, reflecting the influence of socio-economic, demographic and cultural attributes (22,197,198). Public health surveillance to identify neighborhoods at heightened risk of nutrition-related chronic diseases is necessary to develop targeted interventions addressing local barriers to healthy eating (21) and relies on the capacity to monitor dietary patterns within small areas. However, surveys of dietary habits fail to provide an adequate statistical precision at a small spatial scale due to limited sample sizes. In Canada, stable local estimates can only be achieved from existing dietary surveys at the provincial level (24). Without small area measures of dietary behaviors, geographic surveillance of chronic disease risks has used indicators of the built food environment as proxies for diet (199–201). These indicators, known as community food (or nutrition) environment measures, represent the geographic availability of (un)healthy food outlets and are readily computed at small spatial scale from widely available commercial or government retail business registries (202). However, these measures capture the influence of the built environment on dietary behavior rather than actual dietary patterns.

Because most grocery shopping for healthy and unhealthy food occurs in chain supermarkets (97), analyzing sales in these stores can provide insights into what people buy in surrounding communities. Electronic Point-Of-Sales (POS, i.e. store-level rather than for individual shoppers) transaction data are continuously collected from a geographically representative sample of geocoded chain stores including supermarkets, pharmacies and supercenters. We propose to use these data to identify communities at higher risk of nutrition-related chronic diseases.

Our primary objective was to demonstrate a spatial model-based approach to generate a surveillance indicator from POS data of food demand for small areas. Although over 100 food categories are recorded, as an initial example of the methodology, we focused on soda since its consumption is associated with obesity and T2D (181) and is inversely associated with fruit and vegetable intake (203) and meeting dietary guidelines (204).

To be useful, the indicator should estimate the *potential demand* for soda in all areas, including areas with no stores. Demand can be estimated indirectly in areas without stores using model-based small area estimation to account for residents shopping in neighboring areas with stores. Briefly, we decompose the store-level sales of soda as the sum of store- and area-level characteristics including a latent spatial random effect, which captures the variation in the outcome that remains after adjusting for the store- and area-level covariates. Following the Bayesian paradigm, this latent random effect is estimable for areas without any stores through spatial interpolation (i.e. “borrowing information” across neighbors). Our indicator of soda purchasing is this random effect representing *the latent demand (estimated area-level residual sales) of soda, rather than the direct measure of soda sales* that are not observed in all small areas due to the absence of stores in some areas.

Our secondary objective is to estimate the extent to which our indicator is associated with the prevalence of T2D using a disease mapping model, a frequently used public health tool to assess geographic distribution of disease outcomes based on area-level covariates (23).

5.5 Methods

This was an ecological, cross-sectional study that developed an area-level indicator of latent soda demand using POS transaction data and investigated its association with T2D prevalence in 2012. The target region was the Census Metropolitan Area (CMA) of Montreal, which included a population of approximately 3,824, 000 inhabitants (139).

5.5.1 Transaction Data and Store Attributes

The POS data, sold commercially as MarketTrack (Nielsen, New York City, U.S.A), are available in many countries and were purchased from a global marketing research company, Nielsen Corporation (176). To create a representative sample of *chain* retail supermarkets, pharmacies, and supercenters (e.g. Wal-Mart), the company performed stratified random sampling, with the strata defined by store size and store type, namely pharmacies, supermarkets, and supercenters. The transaction data, along with store location and unique identification for each retail chain were transferred electronically to Nielsen from sampled stores, with periodic in-store auditing to ensure reproducibility (205). Based on gross sales from all food outlets from a comprehensive business registry database (described below), the proportion of overall grocery sales of our sampling frame (chain stores) among all chain and independent stores selling food items was 65.2% of the Montreal CMA. Convenience stores were not included in the analysis, since the exact locations of sampled convenience stores were not available, potentially leading to misclassification. Transaction data for warehouse stores (i.e. Costco) are not available from the Nielsen Company in Canada. Data from independent (non-chain) stores and fast food restaurants were not collected by the company.

There were 1,097 chain retail stores (125 sampled chain retail stores and the remaining 972 chain retail stores that were not sampled, or "out-of-sample" stores) in 2012 in the CMA of Montreal. Although POS data were not available for the out-of-sample stores, their locations and chain names were available from Canadian Business Points (Pitney Bowes Canada, Mississauga, Canada) , a commercial business registry that provides an annually updated (including 2012) list of operating businesses including the name, location, store type as defined by the North American Industry Classification code (206), and business size as the number of employees (147). A field validation showed a strong correlation of business points listed in Canadian Business Points with manually verified existing supermarkets, convenience stores and fast-food restaurants (148).

5.5.2 Spatial Unit of Analysis

The CMA of Montreal was partitioned into areas defined by the public health agency of Montreal together with local communities to maximize the homogeneity of socio-demographic and economic attributes (144). This spatial scale in terms of areal size (median: 5.72 km^2 [Interquartile Range (IQR): 2.52 – 18.23 km^2]) and the number of residents (median: 18,410 residents [IQR: 11,870 – 25,530 residents]) is relatively small. The location of the sampled and out-of-sample stores and the areal boundary of the Montreal CMA are shown in Figure 5-1. Because census data were not available for First Nations communities (dotted areas in Figure 5-1), these areas were excluded from the study. Of the 193 areas included, 17 areas had no sampled or out-of-sample chain stores, and 93 areas had chain stores but did not contain sampled chain stores (Supplementary Figure 5-1). Among the 176 areas containing chain stores, the mean number of sampled chain stores in each area was 0.7 (IQR: 0-1.0) and the mean number of both sampled and out-of-sample chain store was 6.2 (IQR: 3.0-8.2).

5.5.3 Statistical Analysis

The analysis followed two steps. First, we modeled the mean of store-level sales of soda as a function of store and area-level covariates and an area-level random effect. The spatial random effect captures the contribution of spatially structured, unmeasured (latent), area-level factors associated with soda sales after accounting for observed store and area-level attributes as covariates. This is our area-level indicator of latent demand of soda and was estimated for all areas, including those that did not contain any chain stores. We then included this indicator as a covariate in a disease-mapping model to estimate area-level relative risk of T2D prevalence along with traditionally used area-level covariates known to be associated with the incidence and prevalence of T2D. We then compared model fit with and without our indicator.

Soda sales model.

The outcome was the store-level proportion of soda sales over all beverages (including diet soda) in 2012 that was logit-transformed to follow a normal distribution. The sales quantity was standardized using the Food and Drug Administration standard serving size of 240ml. The transaction data for the 972 out-of-sample chain stores were estimated through the posterior predictive distribution in the Bayesian paradigm. We assumed that the sales outcome at store j located in area i (D_{ij}) followed a hierarchical normal distribution with mean μ_{ij} and variance σ^2. We decomposed the mean as the sum of store and area-level characteristics by letting,

$$\mu_{ij} = \beta_0 + \beta X_{ij} + \phi_{c[j]} + Z_i \ ,$$

where X_{ij} represents the size of store j in area i measured as the number of employees of the store. The component $\phi_{c[j]}$ is a random effect capturing the association of soda sales with store chain c to which store j belongs. Unlike a dummy variable in a fixed effect model, there is no baseline category to which the indicator of store chain is compared.

Area-level attributes in the model are captured by Z_i, and are the sum of two components, $Z_i = A_i\varphi + S_i$, where A_i is a vector with characteristics of the area containing the store, φ is a vector of the corresponding coefficients , and S_i is a latent spatial random effect representing the remaining area-level variation in soda demand after accounting for the store and area-level characteristics. Specifically, A_i includes area-level median family income, education as the proportion of individuals over age of 25 with post-graduate certificate or diploma, population density as the number of residents per square kilometer, proportion of residents under 18-years old, family size as the mean number of family members, and employment rate as the proportion of those in the labour force employed in full- or part-time work, obtained from the 2011 National Household Survey (139). All the covariates were standardized to have a zero mean and standard deviation of one.

The spatial random effect $S = (S_1 \cdots, S_n)^T$ captures the spatial dependency of soda demand across areas using a proper Conditional Autoregressive (CAR) prior (157), that is,

$$S_i | S_k, k \in \delta_i \sim N\left(\rho \sum_{k \in \delta i} \frac{S_k}{n_i}, \frac{\tau^2}{n_i}\right),$$

where δ_i represents the index set of neighboring areas of i; we assumed that areas i and k were neighbors if they shared a common boundary. The parameter ρ captured the strength of the spatial association among the components of S. In short, the conditional mean of S_i is the mean of the random effects in neighboring areas and controlled by a scaling factor ρ, and the conditional variance is proportional to the number of neighboring areas, with variance parameter τ^2. As the prior of S borrows information of sales across areas, we were able to estimate its value even for areas containing no chain stores. Sampling from the posterior distribution of the parameter vector was done using Markov chain Monte Carlo through OpenBUGS software (207). Further description of the Markov chain Monte Carlo and prior specification is provided in Supplementary Appendix 1.

Diabetes prevalence model

T2D cases were identified using hospital discharge abstracts and physician claims obtained from a 25% random sample of individuals from the Regie de l'Assurance Maladie du Quebec, the public health insurance registry covering 99% of the population in Quebec. Residents receiving at least one hospital diagnostic code (International Classification of Diseases version 10) of T2D or at least two diagnoses of T2D in the physician claims database (International Classification of Diseases version 9) within a two-year period were defined as cases of T2D. The cases in 2012 were geocoded at the level of Canadian postal code and were aggregated to the neighborhood level (n=193). Exclusion criteria were gestational diabetes and under the age of 2 for both cases and non-cases. A previous validation study estimated that the sensitivity and specificity of the T2D ascertainment algorithm were 82.3% and 96.9%, respectively (208). As a sensitivity analysis, we re-fit the diabetes model to adjusted counts of T2D cases generated by a published correction factor (Supplementary Appendix 2).

Let Y_i be the prevalent count of T2D in area *i* in 2012. We assume that Y_i follows a Poisson distribution with mean $e_i\theta_i$, where e_i is an offset term such that θ_i is a relative risk of T2D in area i. Specifically, e_i is the age-adjusted expected number of cases in each area calculated as the product of the number of at-risk individuals in area i (i.e. indirect standardization) and the baseline prevalence, which is the global proportion of cases over at-risk individuals in the entire CMA in 2012 (Supplementary Appendix 3). The component θ_i is modeled as

$$log(\theta_i) = \alpha + G_i\gamma + M_i\eta + S_i\xi + \epsilon_i \ ,$$

where α is an intercept, G_i is a vector containing the area-level covariates: income and education as in the sales model, proportion of immigrants, and a measure of built environment representing the environmental (contextual) influence on physical exercise, which is the number of recreational and fitness facilities per 1,000 residents. In the absence of a measure of physical exercise at a small spatial scale, we were limited to using this measure, which others have used as a potential environmental driver of physical exercise (81) along with similar indicators, such as the availability of greenspace and the local walkability score.

To capture area-level dietary patterns, we compared two indices: a built environmental indicator of store composition known as the modified Retail Food Environment Index (mRFEI) denoted as M_i and our area-level indicator of latent soda demand as component S_i. The mRFEI is an example of a 'community food environment index', calculated as the ratio of the number of healthy food retailers (i.e. stores providing a range of healthy food including fresh fruits and vegetables) to that of unhealthy ones. Healthy food outlets include supermarkets, fruits and vegetable stores, and warehouse stores, while unhealthy ones include convenience stores, fast food restaurants, and small (independent) grocery stores that are defined according to the North American Industry Classification System code (89). Given the predominant assortment of packaged food and the general lack of fresh fruits and vegetables, we classified pharmacies as 'unhealthy'.

The mRFEI is commonly used as a metric of 'food deserts' (mRFEI of zero; areas lacking access to stores providing healthy food) and 'food swamps' (low mRFEI; areas with an over-abundance of unhealthy food sources) (202). Although mRFEI and other community food environment measures represent environmental influence on food choices rather than actual dietary patterns, they are used as a proxy for dietary behaviors in geographic surveillance of disease outcomes (199–201) in the absence of direct behavioral measures. Note that mRFEI cannot be calculated in areas that do not contain any stores. Thus, in the one area that did not contain any stores, mRFEI was assigned a value of zero (designated as 'food desert'). We did not assign a missing value and exclude this area from our analysis, since the area was likely to be at a high-risk of unhealthy diet and chronic disease (given that measures of income and education were in the first quartile), and therefore of interest for surveillance.

Our area-level indicator of latent soda demand (S_i) is available for all geographical areas, regardless of the presence of stores. All the covariates were rescaled to a zero-mean and a standard deviation of three. The number of recreational and fitness facilities and mRFEI were calculated from the Canadian Business Points data as described in Supplementary Appendix 4.

Similar to S_i (our indicator) in the sales model, the component ϵ_i is a spatial random effect commonly used in disease mapping models to capture the spatially and non-spatially patterned structure remaining after accounting for the covariates in the T2D model, while allowing for over-dispersion in the data (209). For this random effect $\epsilon = (\epsilon_1 \ldots \epsilon_n)^T$, we assign a CAR prior (151) assuming a binary neighborhood structure as in the sales model where areas that share boundaries are neighbors. Because the area-level indicator, S_i, already contains a built-in proper CAR structure from the sales model, including a spatially smoothed random effect ϵ_i with a CAR prior in the diabetes model may induce collinearity with S_i, an increasingly recognized phenomenon known as spatial confounding (154,210). To avoid this, we implemented a restricted spatial regression as developed by Hughes and Haran (156) and implemented in the ngspatial package in R software (211).

To evaluate the improvement of model fit with the inclusion of our indicator as compared to the model including the previously used community food environment index (i.e. mRFEI) in disease mapping, we used the deviance information criterion, where a smaller deviance information criterion indicates a superior model fit, and thus more accurate estimation of area-level relative risk of diabetes. We do not present a map of re-classified area as ‘high-risk’ upon addition of our indicator, since the reference (true) risk surface is not available. To explore the potential impact of excluding sales from convenience stores, we performed a sensitivity analysis by adding the area-level number of convenience stores per resident to the disease mapping model containing our indicator.

5.6 Results

Descriptive analysis

Table 5-1 shows the characteristics of the sampled stores. The mean weekly sales of soda varied widely across stores, both in absolute quantity and as a proportion of all beverage sales, reflecting the type of retail format (e.g. superstore and pharmacy) and chain, which differ in size and in the number of available items. The distributions of Socio-Economic Status (SES) and demographic attributes in areas containing the sampled stores (Table 5-2) were comparable to the distributions for all areas in the CMA of Montreal (Table 5-3). Thus, the store sample was representative in terms of reflecting neighborhood characteristics.

Sales model

The association of soda sales with the area-level SES and demographic factors was largely inconclusive (95% credible intervals crossing zero), with the exception of education, which had a negative association with the proportion of soda sales (Figure 5-2, Supplementary

Appendix 5, and Supplementary Table 5-1). In contrast, the indicators of store chain had a substantially stronger association with sales, suggesting that the proportion of soda sales was largely driven by chain-level factors. Comparison of the fitted and observed value of the sales demonstrated good model fit (Supplementary Figure 5-2). Mapping the indicator of latent soda demand (Figure 5-3A) reveals a relatively uniform spatial distribution of the indicator on the island of Montreal (the large island in the center), but a slightly higher demand in the eastern part of the CMA. A map of the posterior standard deviation of our indicator (Figure 5-3B) shows higher values in areas containing a smaller number of sampled stores (Supplementary Figure 5-3A) or having a smaller number of neighbors (Supplementary Figure 5-3B). This is expected, since there is less information to learn from a smaller number of observations (i.e. sampled stores), and the conditional variance of the CAR prior is inversely proportional to the number of neighbors.

Type 2 Diabetes model

After adjustment for area-level covariates, addition of the indicator of latent soda demand to the diabetes model improved model fit relative to a disease-mapping model including the mRFEI (deviance information criterion = 2140 with the soda demand indicator vs. 2148 with mRFEI); an 8 unit improvement, which is substantial (212) (Table 5-4). All covariates were associated with area-level T2D prevalence (i.e. 95% credible interval excluding 1). Figure 5-4 shows the spatial distribution of high, low and indeterminate risk area classified based on the posterior credible interval of the relative risk estimated by the model containing our area-level indicator of latent soda demand (the map showing the point estimates, or posterior mean, of the relative risk is provided in Supplementary Figure 5-4). An increased risk of diabetes is evident in the east of the main island of Montreal.

Sensitivity analysis accounting for potential misclassification of T2D gave results similar to our main analysis for the regression coefficients and a difference of 6 units in deviance information criterion, which is still substantial (Supplementary Table 5-2). A sensitivity analysis to explore the impact of not including sales from convenience stores found that the number of convenience stores per resident was not a significant predictor of T2D prevalence (posterior mean: 0.99, 95% credible interval: 0.95-1.03), and inclusion of this variable altered neither the posterior mean nor the credible interval of our indicator.

5.7 Discussion

We developed an area-level indicator of the latent demand for soda using POS transaction data in the CMA of Montreal, Canada. Our indicator for chronic disease surveillance at small spatial scale can be estimated for all geographical areas and improved the fit of a disease mapping model for estimating the area-level relative risk of T2D. The indicator represents the latent residual demand for soda sales, and its positive association with the relative risk suggests that unobserved spatially correlated drivers of soda sales also had a positive relationship with the area-level T2D risk. The use of a spatial random effect allowed estimation of demand in 17 areas without stores by smoothing information from neighbors and accounting for the mobility of shoppers purchasing soda in adjacent areas.

The improvement of model fit with our indicator over mRFEI is substantial. This result is not surprising as, conceptually, our indicator is generated from actual soda sales and thus is a more appropriate measure of the geographic risk of diet-related chronic diseases as it describes the behavior of individuals in an area. Built environment measures such as mRFEI are useful to identify "food desert" and "food swamp" areas in food environment research and in surveillance of the built environment. However, the use of these measures as a surrogate for local dietary patterns in the surveillance of geographic disease risk is limited because dichotomization of stores into ‘healthy’ and ‘unhealthy’ fails to acknowledge varying sales across stores. Our results adds support to a growing recognition that the neighborhood presence of ‘healthy’ stores does not equate with healthy purchasing patterns (213). In addition, we observed one area in Montreal with neither chain nor independent stores, for which the mRFEI was undefined (we used ad-hoc approach of imputing it with zero).

The map of T2D risk showed a West-to-East spatial trend across the island of Montreal, which is similar to the spatial patterning reported previously for the crude prevalence of other chronic diseases and material deprivation (214). The negative associations between T2D risk and education, median family income, and recreational facilities are consistent with previous findings (199,215,216).

.

Marketing researchers have used POS transaction data extensively for the past three decades to monitor consumer behaviors and to develop and evaluate marketing programs (27). In contrast, public health researchers have limited use of these data to monitor dietary patterns and nutrition-related chronic diseases, other than to analyze purchasing over time (217). To our knowledge, this is the first study to perform spatial statistical modeling of POS data from a public health perspective. National-level dietary interventions are often poorly followed, emphasizing the importance of local monitoring to inform more precise actions tailored to community-specific characteristics (21,22). Use of our indicator should improve local disease risk assessment and aid in the allocation of public health resources to areas where the greatest benefit can be expected, while allowing the assessment of local response to these interventions.

Limitations include the use of a single food category. Although soda consumption is associated with dietary patterns inconsistent with nutritional guidelines, future application of our methods to multiple categories may better describe ‘unhealthy’ purchasing patterns, potentially allowing more precise estimation of the risk of nutrition-related chronic diseases. As well, because our store sample did not include convenience stores, warehouses and restaurants, the observed sales were not a complete measurement of soda purchasing. Nevertheless, our models included the most important store type, supermarket, which is a major venue for food purchasing due to their large store size and the affordability of the food sold (97,190). Finally, our findings apply to associations among area-level measures, and caution should be exercised in interpreting these findings in terms of individual-level soda consumption and T2D.

Future research includes evaluating the utility of our indicator of latent food demand for predicting other nutrition-related health outcomes such as obesity, a major manifestation of excess calorie intake and an important risk factor for many chronic diseases (218). Given the automated collection of transactional data, another research opportunity is to assess the temporal association of our indicator and chronic diseases over many years.

Health and non-health related electronic data are expected to play an increasingly important role in public health surveillance (219), yet measurement of population diet at high spatial resolution is lacking. We have proposed a spatial modeling approach that uses widely-available POS transaction data to develop an indicator of the latent demand of soda at a small spatial scale. Our findings suggest that purchasing data for soda can improve the accuracy with which the geographic risk of T2D is estimated, thus building a foundation for further exploration of these data for the measurement of local dietary patterns.

Table 5-1. Characteristics of Sampled Stores (N=125) in Census Metropolitan Area of Montreal, 2012

Store characteristics	Min	Mean	Median (IQR)	Max
Weekly soda sales in servings [a]	8.4	18,509.3	13,071.9 (1,322.6-26,964.6	108,771.7
% of weekly soda sales over all beverage sales [a]	0.7	16.1	15.4 (12.2-19.4)	41.4
Number of full-time employed individuals	2.0	35.6	9.5 (5.0-50.0)	270.0

Abbreviations: IQR, interquartile range; Max, maximum; Min, minimum

[a] These quantities are average of weekly beverage transaction in 2012, calculated from 125 sampled stores.

Table 5-2. Characteristics of Neighborhoods on Which Sampled Chain Stores Were Located (83 areas) in Census Metropolitan Area of Montreal, 2011 Canadian National Household Survey

Neighborhood characteristics	Min	Mean	Median (IQR)	Max
Education [a]	29.6	68.0	67.6 (61.5-74.8)	90.3
Median family income (in 10,000 Canadian dollars)	2.0	7.4	7.0 (5.8-8.7)	16.7
% of immigrant	1.6	21.6	20.4 (8.1-32.6)	63.1
% of residents under 18 years old	2.4	19.7	20.5 (17.2-23.3)	32.9
% of employed residents among labor force	18.5	59.9	59.0 (54.1-66.9)	76.8
Average family size	1.1	2.9	3.0 (2.9-3.1)	3.5
Population density (residents per square kilometer)	109.2	3,811.8	2,298.6 (1,074.1-	19,606.1

Abbreviations: IQR, interquartile range; Max, maximum; Min, minimum

[a] % of residents with post-graduate diploma or certificate among age greater than 25

Table 5-3. Characteristics of Neighborhoods (193 areas) in Census Metropolitan Area of Montreal, 2011 Canadian National Household Survey

Neighborhood characteristics	**Min**	**Mean**	**Median (IQR)**	**Max**
Count of T2D	43.0	346.1	319.0 (211.0 - 436.0)	1617.0
Prevalence of T2D (%)	3.1	7.0	6.8 (5.8-8.2)	10.7
Education [a]	29.6	67.1	66.3 (60.5-74.8)	90.3
Median family income (in 10,000 Canadian dollars)	2.0	7.2	6.9 (5.5-8.5)	16.7
% of immigrant	1.6	22.6	20.9 (9.5-32.7)	63.6
% of residents under 18 years old	2.4	19.9	20.9 (17.5-23.3)	32.9
% of employed residents among labor force	18.5	59.0	59.0 (54.1-66.3)	76.8
Average family size	1.1	2.9	3 .0 (2.8-3.1.0)	3.5
Population density (residents per square kilometer)	54.7	4,023.5	3,348 (1,093.5-5,744.8)	19,606.1
modified Retail Food Environment Index (mRFEI)	0.0	15.8	14.9 (10.5-20.0)	75.0
Number of recreational facilities per 1,000 residents	0.0	0.4	0.3 (0.2-0.5)	2.8

Abbreviations: IQR, interquartile range; Max, maximum; Min, minimum; T2D, Type 2 Diabetes Mellitus

[a] % of residents with post-graduate diploma or certificate among age greater than 25

Table 5-4. Deviance Information Criterion, Posterior Mean and 95% Credible Interval of Exponentiated Coefficients in Diabetes Risk Model, Census Metropolitan Area of Montreal, 2012

Regression coefficients	**Reference model**		**mRFEI**		***S***	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Intercept	0.99	0.98, 1.00	0.99	0.98, 1.00	0.99	0.98, 1.00
Education [a]	0.74	0.71, 0.77	0.75	0.72, 0.78	0.74	0.71, 0.77
Immigrant [a]	1.16	1.13, 1.20	1.16	1.12, 1.19	1.16	1.12, 1.20
Income [a]	0.84	0.80, 0.88	0.83	0.80, 0.87	0.84	0.81, 0.88
Recreational facility [a]	0.94	0.91, 0.98	0.95	0.92, 0.99	0.94	0.91, 0.98
mRFEI [a]			0.94	0.91, 0.97		
Indicator of latent soda demand (*S*) [a]					1.06	1.03, 1.08
DIC	**2,157**		**2,148**		**2,140**	

Abbreviations: 95% CI, 95% credible interval; DIC, deviance information criterion; Mean, posterior mean; mRFEI, modified Retail Food Environment Index

[a] Covariates were mean centered and scaled to three standard deviations, and the regression coefficients were exponentiated. The value of coefficients represents an area-level relative risk of T2D, which is the ratio of the risk at one unit increase of the covariates to the risk at mean value of the covariates.

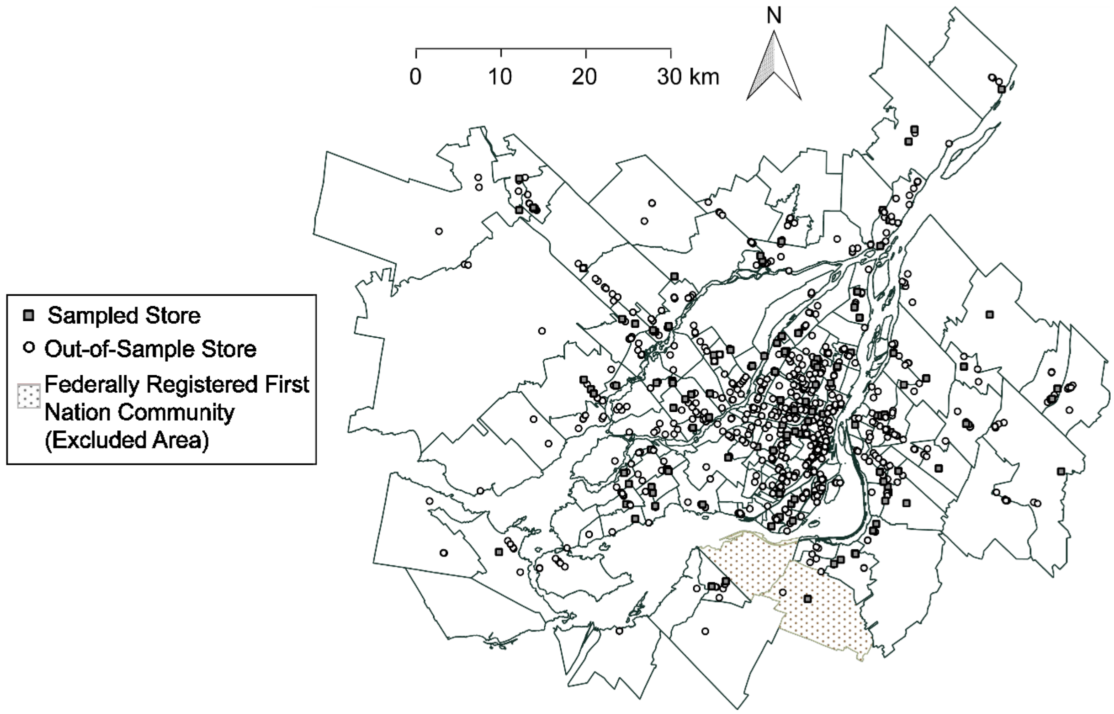


Figure 5-1. Location of Sampled and Out-of-Sample Chain Retail Supermarkets, Pharmacies, and Supercenters Placed in 193 Areas in the Census Metropolitan Area of Montreal, 2012

Black square indicates sampled store, and empty circle indicates out-of sample store. Dotted areas are the federally registered First Nations communities (excluded areas). Solid Black borders represent areal boundaries.

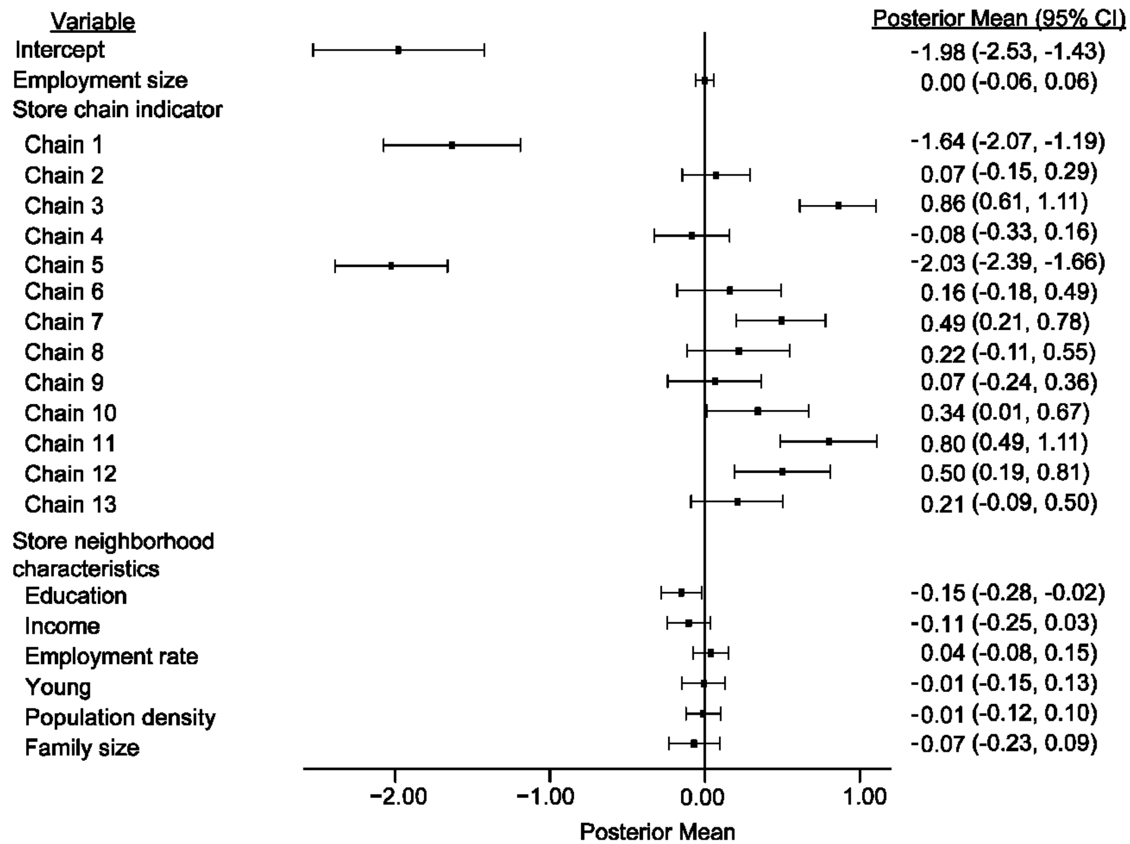


Figure 5-2. Estimated Association of Store and Area-level Covariates with the Logit-Transformed Proportion of Store-Level Soda Sales, Census Metropolitan Area of Montreal, 2012

Employment size indicates the number of full-time employees in each store.

Store chain indicator represents retail chain identifier for random effect. Unlike a dummy variable in a fixed effect model, there is no baseline category to which the indicator of store chain is compared. Chain 1-5 are pharmacy chain, chain 6-11 are supermarket chain, and chain 12 and 13 are supercenter chain.

Store neighborhood characteristics consist of continuous variables standardized to have mean zero and standard deviation of one. The covariate "Young" represents the proportion of residents under 18 years old, and the covariate "Family size" represents the mean number of family member.

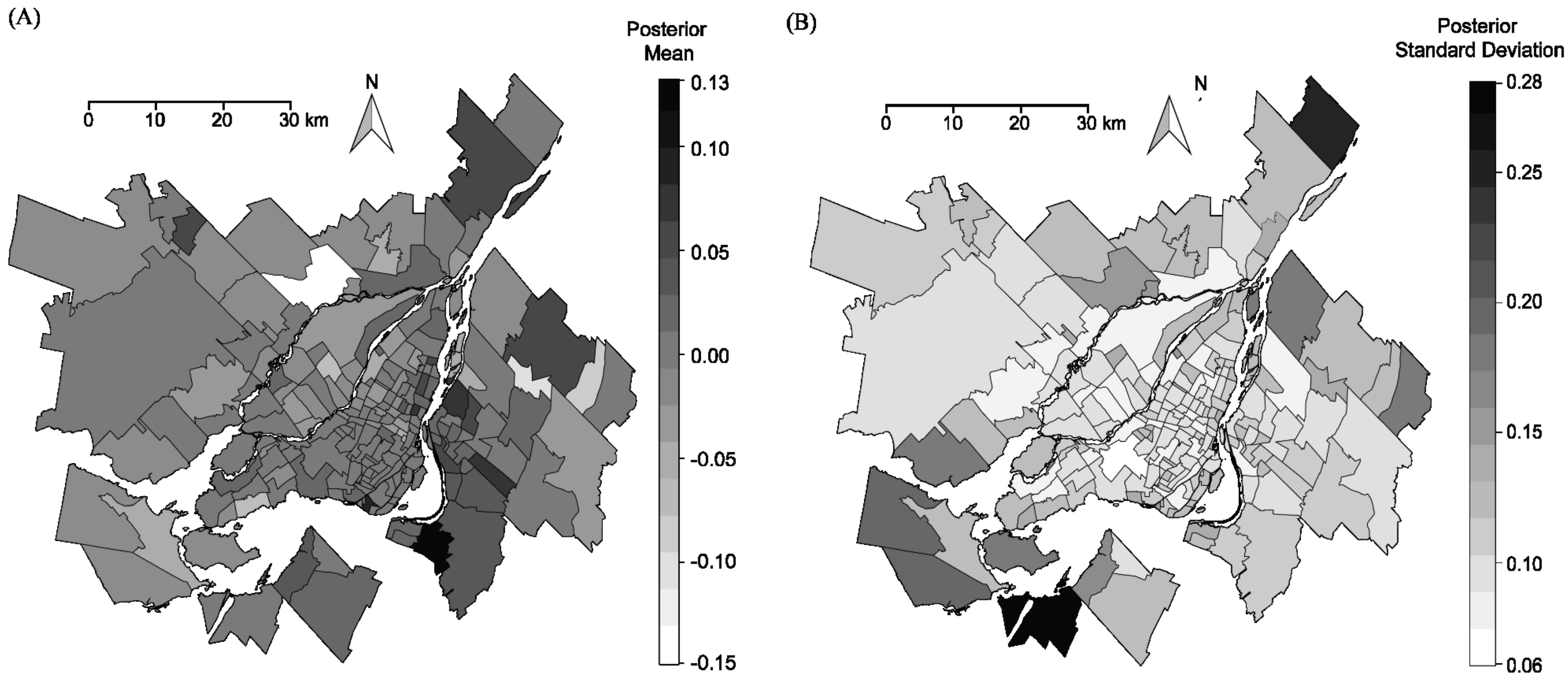


Figure 5-3 A) and B). Posterior Mean A) and Standard Deviation B) of Area-Level Indicator of Latent Soda Demand, Census Metropolitan Area of Montreal, 2012

The grey key on the right of each figure indicates the value of the posterior mean A) and standard deviation B), respectively.

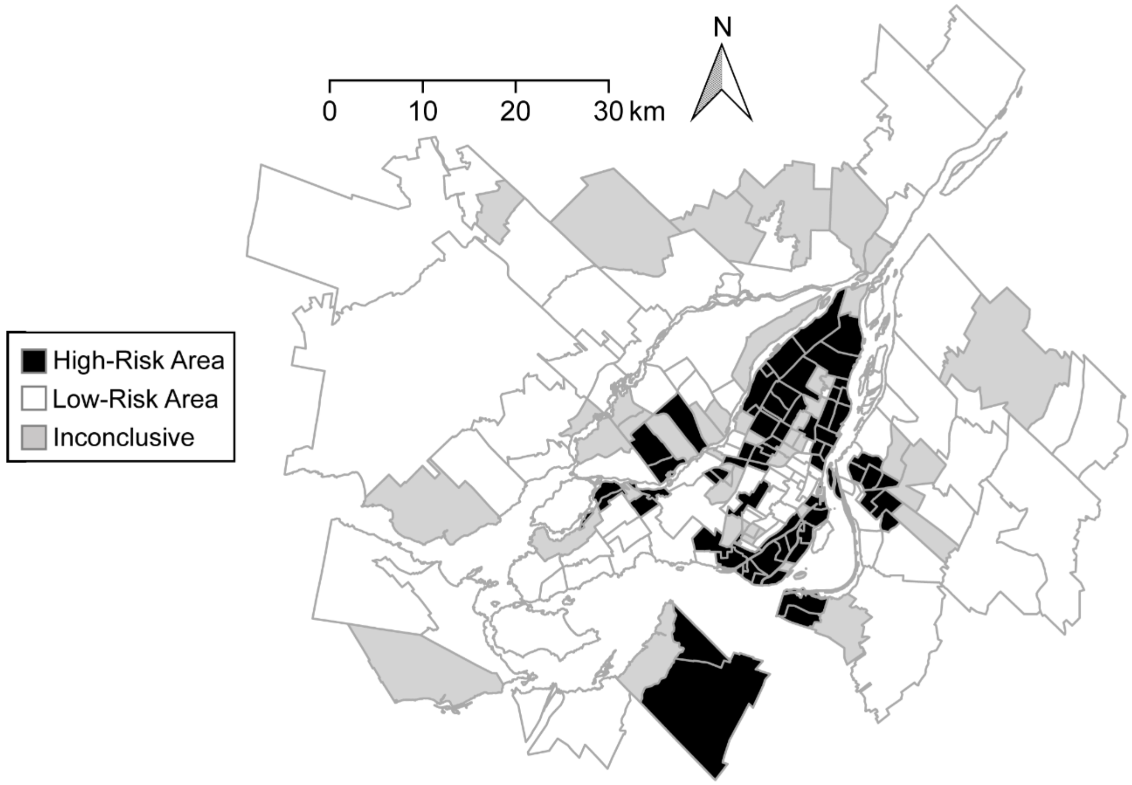


Figure 5-4. Classification of Areas by Relative Risk of Type 2 Diabetes Mellitus as Estimated by the Model Containing Areal-Level Indicator of Latent Soda Demand, Census Metropolitan Area of Montreal, 2012

Black fill indicates high-risk area where lower (2.5%) boundary of the credible interval of the relative risk is greater than one. White fill indicates low-risk area where upper (97.5%) boundary of the credible interval is less than one. Grey fill indicates area with inconclusive classification (95% credible interval contained the relative risk of one).

5.8 Supplementary Materials for Manuscript 2

Appendix 1. Prior Specification and Markov Chain Monte Carlo (MCMC) of the Soda Sales and Diabetes Model

Prior distribution selected for regression parameters, φ, β_0 and β was $Normal(0,100)$, and the prior distribution for precision parameters, σ_ϕ^{-2}, σ^{-2}, and τ^{-2} was $Gamma(0.5,0.01)$. The range of the scaling parameter ρ was determined as the inverse of the minimum and maximum eigenvalues of $D_w^{-1/2} W D_w^{-1/2}$, where W was the normalized adjacency weight matrix, and D_w represents a diagonal matrix whose *i* th diagonal element is equal to the number of neighbor (1). To improve the mixing of MCMC, we implemented hierarchical centering (2), such that the prior mean of the chain random effect is centered at the overall intercept, that is, $\phi_{cj} \sim N(\beta_0, \sigma_\phi^2)$.

We let the MCMC run for 60,000 iterations, considered the first 10,000 as burn-in and stored every 20th iteration to avoid autocorrelation among the sampled values. Monte Carlo Standard Error was less than 5% of sample deviation. Trace plots of three MCMC chains starting with very different initial values were visually inspected to investigate convergence.

For the restricted spatial regression (diabetes model) implemented in the ngspatial package, default prior distributions recommended by Hughes and Haran (3) were used. The number of attractive Moran eigenvector was specified as 50, and 2 million iterations was performed. As in the soda sales model, convergence was assessed by visual inspection of MCMC chains.

Appendix 2. Sensitivity Analysis Adjusting for Potential Misclassification of Type 2 Diabetes Mellitus (T2D) Diagnostic Algorithm.

Using a correction factor of diabetes prevalence from a meta-analysis of previous validation research (4), we re-calculated area-level prevalence and expected count and re-fit the diabetes risk model. The results (Supplementary Table 5-2) indicate that the association of all covariates with T2D risk increased very slightly (i.e. away from the null value of 1.0).

Appendix 3. Calculation of age-Adjusted Expected Count of Diabetes.

Age-adjusted expected count for area i was calculated by indirect standardization;

$e_i = \sum_{k=1}^{8} N_{ik}\, p_k$,

where N_{ik} is the population size greater or equal to age two in area i at age stratum k.

Age strata k were defined as [2-20) [20-30) [30-40) [40-50) [50-60) [60-70) [70-80), and>=80 years old. Background (reference) disease prevalence in age stratum k in the study region (CMA of Montreal) was denoted as p_k and calculated as the number of people having diabetes status in 2012 in stratum k divided by the corresponding age stratum-specific total population with and without diabetes diagnosis.

Appendix 4. Calculation of Area-Level Retail Food Environment Index and Recreational Facilities.

Healthiness of retail food environment was created by the modified Retail Food Environment Index (mRFEI), a measure created by the U.S Centers for Disease Control and Prevention for the census tract-level scanning of (un)healthy food environment (5). The measure is defined as;

$$\text{mRFEI} = \frac{Number of healthy food retailers}{Number of healthy + unhealthy food retailers}$$

in each area. Using the North American Industry Classification System code (NAICS) assigned to food outlets in the Canadian Business Points business listing, we classified healthy food retailer as stores having supermarkets (NAICS 445110) and fruit and vegetable markets (NAICS 445230). Less healthy food outlets include fast food restaurants, small (typically non-chain) grocery stores, and convenience stores. Fast food stores were defined according to NAICS code 722211(fast food restaurants). Convenience stores were defined according to NAICS code 445120 (convenience stores) or NAICS code 445110 (small groceries). Although pharmacies are not classified in the original mRFEI, we defined them as unhealthy due to the lack of fruits and vegetables.

Availability of recreational facility was calculated as the number of facilities divided by 1,000 residents for each area. The definition of the recreational facility is based on the Standard Industry Classification (SIC) codes as previously described (6).

Appendix 5. Additional Results of Sales Model.

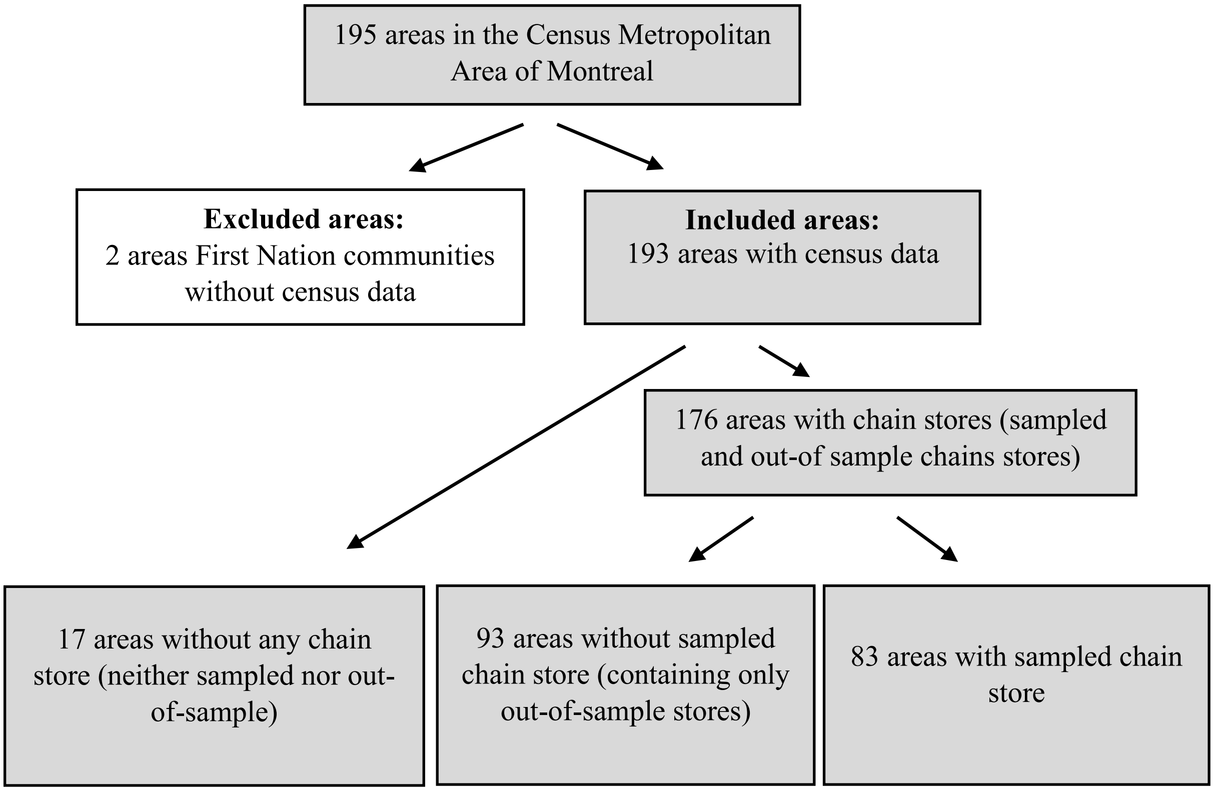
The inconclusive association of area-level income with the proportion soda sales (Supplementary Table 5-1) may be explained by the uniformly affordable nature of soda regardless of household income.

Supplementary Table 5-1. Estimated Association of Store and Area-Level Covariates with the Logit-Transformed Proportion of Store-Level Soda Sales, Montreal CMA, 2012

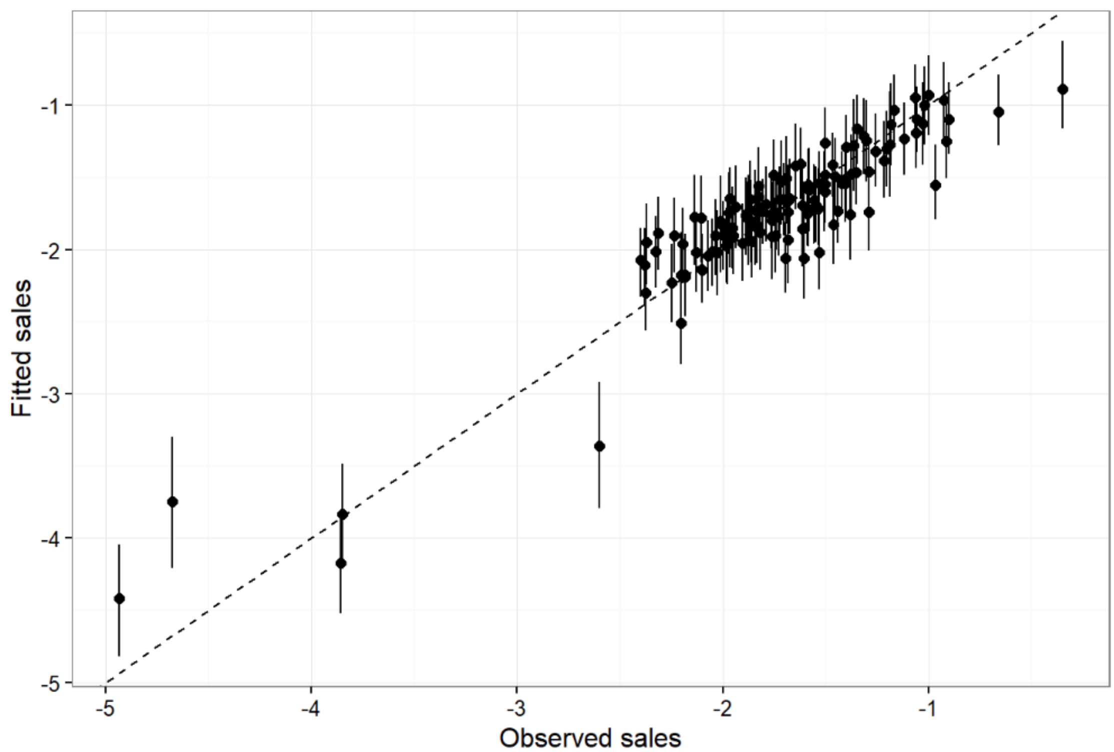
Regression coefficients	**Mean**	**95% CI**
Intercept	-1.98	-2.53, -1.43
Employment size [a]	0.00	-0.06, 0.06
Chain 1[b]	-1.64	-2.07, -1.19
Chain 2 [b]	0.07	-0.15, 0.29
Chain 3 [b]	0.86	0.61, 1.11
Chain 4 [b]	-0.08	-0.33, 0.16
Chain 5 [b]	-2.03	-2.39, -1.66
Chain 6 [c]	0.16	-0.18, 0.49
Chain 7 [c]	0.49	0.21, 0.78
Chain 8 [c]	0.22	-0.11, 0.55
Chain 9 [c]	0.07	-0.24, 0.36
Chain 10 [c]	0.34	0.01, 0.67
Chain 11 [c]	0.80	0.49, 1.11
Chain 12 [d]	0.50	0.19, 0.81
Chain 13 [d]	0.21	-0.09, 0.50
Education [e]	-0.15	-0.28, -0.02
Income [e]	-0.11	-0.25, 0.03
Employment rate [e]	0.04	-0.08, 0.15
Young [e,f]	-0.01	-0.15, 0.13
Population density [e]	-0.01	-0.12, 0.10
Family size [e,g]	-0.07	-0.23, 0.09

Abbreviations: 95% CI, 95% credible interval; Mean, posterior mean

[a] Number of full-time employees in each store, continuous variable
[b] Pharmacy chain
[c] Supermarket chain
[d] Supercenter chain
[e] Area-level covariates; continuous variables standardized to have mean zero and standard deviation of one
[f] Proportion of residents under 18 years old
[g] Mean number of family member

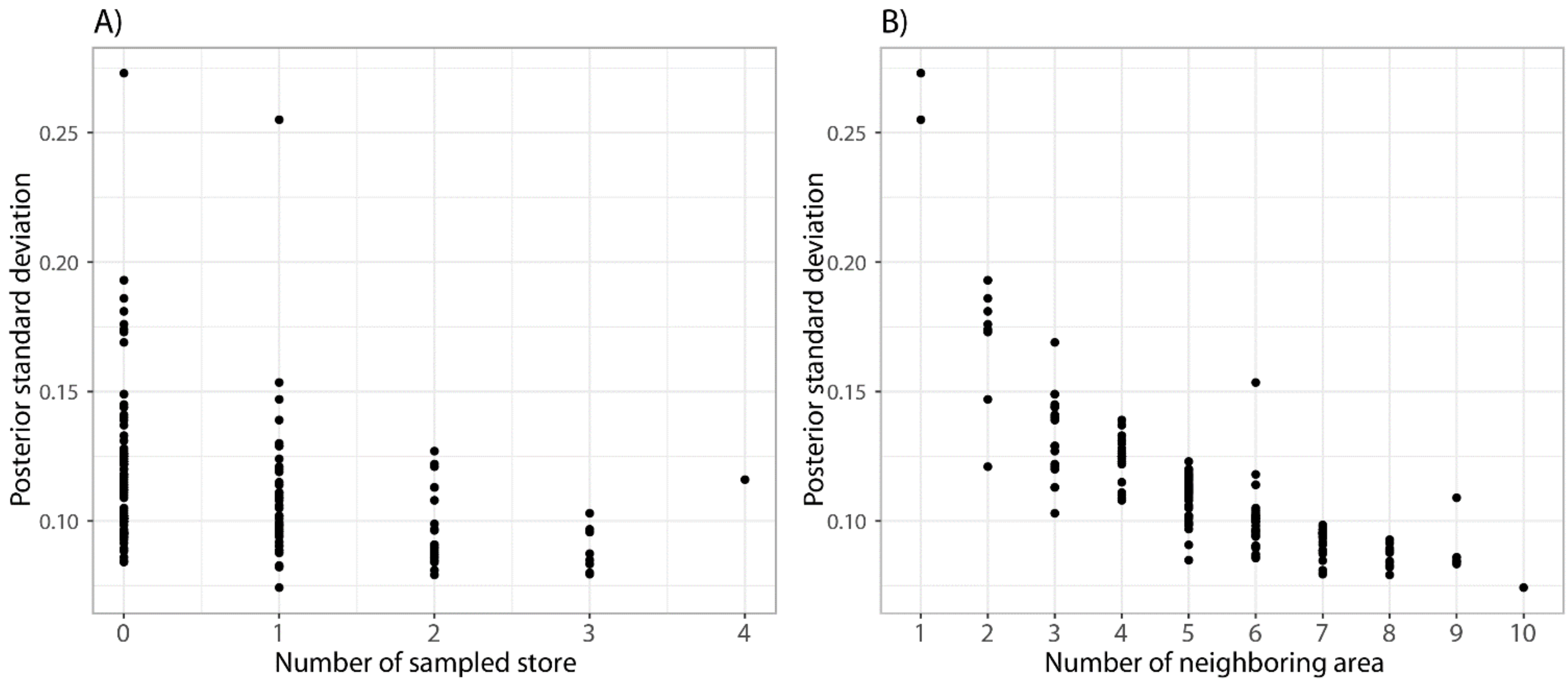


Supplementary Figure 5-1. Store Composition of Areas in Census Metropolitan Area (CMA) of Montreal, 2012

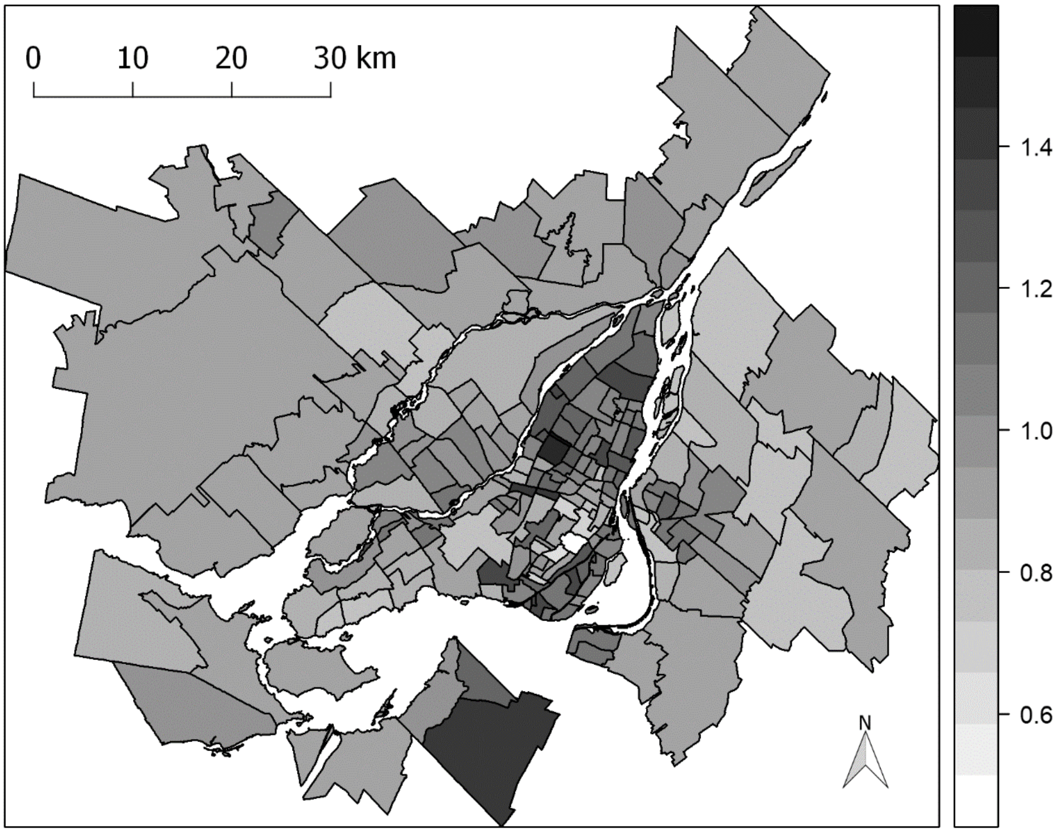


Supplementary Figure 5-2. Posterior Summary of the Observed Versus Fitted Store-Level Proportion of Soda Sales, CMA of Montreal, 2012

Vertical solid line indicates 95% posterior credible interval of fitted logit-transformed proportion of store-level soda sales over other beverages. Dashed line is reference slope, where y = x.



Supplementary Figure 5-3. Relationship of Posterior Standard Deviation of Area-Level Indicator of Latent Soda Demand and A) Number of Sampled Stores and B) Number of Neighboring Areas for Each Area, CMA of Montreal, 2012



Supplementary Figure 5-4. Posterior Mean of Relative Risk of Type 2 Diabetes Mellitus as Estimated by the Model Containing Areal-Level Indicator of Latent Soda Demand, Montreal CMA, 2012

The grey key on the right indicates posterior mean of relative risk for each area.

Supplementary Table 5-2: Deviance Information Criterion, Posterior Mean and 95% Credible Interval of Exponentiated Coefficients in Diabetes Risk Model with misclassification-adjusted T2D prevalence, Montreal CMA, 2012

Regression coefficients	**Reference model**		**mRFEI**		***S***	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Intercept	0.98	0.97, 0.99	0.98	0.97, 0.99	0.98	0.97, 0.99
Education [a]	0.65	0.62, 0.68	0.65	0.62, 0.68	0.66	0.63, 0.69
Immigrant [a]	1.25	1.21, 1.29	1.24	1.20, 1.29	1.24	1.20, 1.28
Income [a]	0.78	0.74, 0.81	0.78	0.74, 0.82	0.77	0.73, 0.81
Recreational	0.91	0.87, 0.94	0.91	0.88, 0.95	0.92	0.89, 0.96
mRFEI [a]			0.91	0.88, 0.95		
Indicator of latent soda demand (*S*) [a]					1.10	1.07, 1.13
DIC	2648		2620		2614	

Abbreviations: 95% CI, 95% credible interval; DIC, deviance information criterion; Mean, posterior mean; mRFEI, modified Retail Food Environment Index

[a] Covariates were mean centered and scaled to three standard deviations, and the regression coefficients were exponentiated. The presented value of coefficients represents an area-level relative risk of T2D, which is the ratio of the risk at one unit increase of the covariates to the risk at mean value of the covariates.

References for Supplementary Materials

1. Banerjee S, Carlin B, Alan G. *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. Boca Raton, FL: CRC Press; 2014.

2. Gelfand AE, Sahu SK, Carlin BP. Efficient Parametrisations for Normal Linear Mixed Models. *Biometrika*. 1995;82(3):479–88.

3. Hughes J, Haran M. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J R Stat Soc Ser B Stat Methodol*. 2013 Jan 1;75(1):139–59.

4. Leong A, Dasgupta K, Bernatsky S, et al. Systematic Review and Meta-Analysis of Validation Studies on a Diabetes Case Definition from Health Administrative Records. *PLOS ONE*. 2013;8(10):e75256.

5. Centers for Disease Control and Prevention. Census Tract Level State Maps of the Modified Retail Food Environment Index (mRFEI). Published 2011 Accessed Feb 29, 2016. ftp://ftp.cdc.gov/pub/Publications/dnpao/census-tract-level-state-maps-mrfei_TAG508.pdf

6. Boone-Heinonen J, Diez-Roux AV, Goff DC, et al. The Neighborhood Energy Balance Equation: Does Neighborhood Food Retail Environment + Physical Activity Environment = Obesity? The CARDIA Study. *PLoS ONE*. 2013 Dec 27;8(12):e85141.

Chapter 6: Generating Area-Level Purchasing From Store-level Sales: Application of a Retail Gravity Model to Create Small-Area Indicators of Food Purchasing in Montreal, Canada (Manuscript 3)

6.1 Preamble

The final manuscript of my thesis makes another methodological contribution to small area estimation of diets using POS data. Distinct from the second manuscript that used spatial smoothing via a CAR model, this manuscript generated an area-level purchasing indicator by partitioning and re-distributing store-level sales into a set of surrounding areas with weights. The weights were origin-destination travel probabilities for the combination of each area (origin) and stores (destination) calculated by a retail gravity model. While the indicator generated in Manuscript 2 is the area-level *latent* spatial effect on soda sales after accounting for the store and area-level predicator of sales, the indicators generated in this manuscript represent area-level purchasing quantity. Unlike Manuscript 1 and 2 that focused on the sales of soda only, this manuscript also included the sales of diet-soda, flavored (sugary) yogurt, and plain (non-sugar) yogurt.

As in Manuscript 2, the utility of area-level purchasing indicator of soda, diet-soda, plain and flavored (sugary) yogurt in the context of spatial disease risk estimation were investigated, specifically for their ability to improve fit of the disease mapping for T2D. Different from Manuscript 2, the disease mapping model in this manuscript included spatial and non-spatial random effects as specified by the BYM convolution prior, with a re-parametrization of the hyperparameters to provide a clearer assessment for the contribution of spatial and non-spatial components on residual disease risk. The modified BYM model showed an improved model fit upon addition of my indicator of the four food categories. With Manuscript 2, this study demonstrates the first spatial analysis POS data in public health research, with the goal of developing area-level consumptions from store-level sales. The application of a retail gravity model to decompose store-level sales into surrounding area-level purchasing provided a feasible solution to generate purchasing indicators in areas lacking stores. This work should stimulate further development of store gravity models as new data become available, for example shopping surveys to learn the mobility parameters. As with Manuscripts 1 and 2, Manuscript 3 opens a methodological journey for the analysis of POS data to overcome existing research and practice gap that are challenging to address with traditional data alone.

The manuscript is in preparation for submission to a peer-reviewed journal.

6.2 Manuscript Title Page

Generating Area-Level Purchasing From Store-Level Sales: Application of a Retail Gravity Model to Create Small-Area Indicators of Food Purchasing in Montreal, Canada.

Hiroshi Mamiya, Alexandra M. Schmidt, Erica E. M. Moodie, Yu Ma, and David L. Buckeridge

Corresponding to Hiroshi Mamiya, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1140 Ave Pine, Montreal, Quebec, Canada, H3A 1A3 (e-mail: hiroshi.mamiya@mail.mcgill.ca)

Author affiliations: Surveillance Lab, McGill Clinical and Health Informatics, McGill University, Montreal, Quebec, Canada (Hiroshi Mamiya and David L. Buckeridge); Department of Epidemiology Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada (Hiroshi Mamiya, Alexandra M. Schmidt, Erica EM. Moodie, David L. Buckeridge); Desautels Faculty of Management, McGill University, Montreal, Quebec, Canada (Yu Ma).

Keywords: Ecological Analysis, Point-of-Sales Transaction Data, Public Health Surveillance, Retail Gravity Model Analysis, Type 2 Diabetes Mellitus

6.3 Abstract

The geographic resolution of dietary surveillance in Canada and other countries is current limited to the national or provincial (state) scale. This limitation is due to a reliance on population health surveys, which have inadequate sample sizes to achieve acceptable statistical precision at higher scales. We developed a small-area indicator of food purchasing for soda, diet-soda, flavored yogurt, and plain yogurt in the Census Metropolitan Area of Montreal in 2012 through the decomposition of store-level grocery transaction data using the Huff retail gravity model. The resulting area-level indicators improved the fit of disease risk mapping model for area-level prevalence of type 2 diabetes mellitus (Watanabe-Akaike Information Criterion = 1769 with the indicators, 1772 without). We demonstrated the application of retail gravity model to enable community surveillance of nutrition risk factors at a high spatial scale.

6.4 Introduction

Unhealthy diet accounts for an estimated 11 million deaths and 255 million disability-adjusted life-years worldwide in 2017 (220). Diet high in sugar is a major risk factor for obesity, a global public health threat leading to numerous Non Communicable Diseases (NCDs) including Type 2 Diabetes mellitus (T2D), cardiovascular disease, and many forms of cancers (5). Risk factors for unhealthy diets exhibit geographic patterning, resulting in regional disparities in diets and spatially heterogeneous burden of NCD burden across city subdivisions (8). These factors, often spatially correlated with each other, include socio-demographic and economic status, walkability, lack of food outlets providing healthful food at an affordable price, and food marketing activities. Interventions including public health policies and community programs are needed to address local barriers of optimal diet in areas at elevated risk of NCDs (2,5–7). Progress tracking and evaluating these interventions require a measurement of dietary patterns at a community level, which is also necessary for the identification of neighbourhoods for targeted intervention (198).

Surveys administering dietary questionnaires have been a major source of nutrition surveillance but are typically unable to measure nutrition status in small administrative districts due to an inadequate number of respondents for a statistically precise estimation (132). An alternative and viable solution to measure community diets is the use of geo-coded Point-Of-Sales (POS, store-level rather than individual consumer) grocery transaction records, which are collected from a sample of retail food stores and represent purchasing activities of residents in surrounding areas. The key methodological challenge of attributing store sales to neighboring areas is mobility. Because shoppers often use food outlets outside their own residential area and do not necessarily shop in stores closest to their residence, the measure of grocery purchasing in one area is likely to be dependent on the sales observed in surrounding areas. We previously performed the first spatial statistical analysis of grocery transaction data in the public health context, , specifically estimating the spatially smoothed demand of soda using a hierarchical Bayesian model (221). Although shown to be predictive of area-level type 2 diabetes mellitus, the estimated values of the indicator may pose a conceptual challenge to interprets to practitioners and decision makers, as they represented 'latent' demand of soda, i.e. residual spatial effects, rather than actual quantity of food purchasing.

As an alternative to spatial smoothing, it is possible to partition store sales into surrounding areas based on the probability of residents in each area visiting surrounding stores. Reilly's law of retail gravitation states that a consumer selects a particular store mainly based upon the store size and the cost of travelling to the store (163). The Huff gravity model is a commonly used and empirically tested probabilistic store catchment area analysis that uses these characteristics to generate a discrete (area-level) surface of the origin-destination probability for all stores in the study region (222). Our primary objective is to develop small area indicators of food purchasing through the decomposition of food sales in the Census Metropolitan Area (CMA) of Montreal using the Huff retail gravity model. Our secondary objective is to evaluate the utility of these indicators in predicting the area-level T2D for disease mapping, an integral tool for public health surveillance in measuring spatial variation and trends of disease risk.

We developed purchasing indicators for soda (carbonated soft drink) and yogurt.

Consumption of soda is a major source of total caloric intake among Canadians and requires ongoing monitoring due to its strong association with T2D and cardiovascular diseases (181,223). Because diet-soda (soda with artificial sweeteners) does not contain free sugar, has inconclusive association with NCDs (224) and associated with distinct consumer characteristics (225), we developed a separate indicator for diet-soda. Solid (spoonable as opposed to drinkable) yogurt products have grown rapidly in North America in terms of consumption, marketing and product diversification (57). Despite increasing evidence for health benefits (61) and wide perception as "healthy", the majority of yogurt products marketed to children contains artificially added sugar to enhance flavor and texture (226), making this food category one of the main sources of total calories among Canadian children in 2015 (53). For this reason, we also developed separate indicators for flavored yogurt and plain yogurt, the latter defined by products with intrinsic (naturally derived from milk) sugars only.

6.5 Methods

This was an ecological cross-sectional study that developed area-level purchasing indicators for the aforementioned target food categories using POS grocery transaction data in the CMA of Montreal, 2012.

6.5.1 Transaction Data

The POS transaction data were purchased from a global marketing company, Nielsen (176), that collects weekly transactions from a stratified random sample of *chain* retail food outlets. There were 1,097 chain food outlets in the CMA of Montreal in 2012 (125 sampled chain stores and the remaining 972 chain stores that were not sampled, or “out-of-sample” stores). These stores consisted of 6 supermarket chains, 5 pharmacy chains and 2 supercenter chains, all selling soda and diet-soda products. Plain and flavored yogurt were sold in a subset of these stores that consisted of 311 stores in total (72 sampled and 239 out-of-sample stores), as they were rarely sold in pharmacy chains and one supercenter chain. The geographic coordinates and name of out-of-sample stores was available from a commercial business directory database, Canada Business Point (CBP) (147). Although transaction of soda and diet-soda from convenience stores were available, we did not include the data from this store type, since the exact locations of sampled convenience stores were not available.

We extracted transaction records of individual food products belonging to the category of soda and spoonable yogurt. We then separated diet soda and plain yogurt according to terms suggestive of diet and plain products and converted the quantity of sales into the Food and Drug Administration standardized serving size. We aggregated the serving-standardised weekly sales of individual food items into the average annual sales in 2012 for the four categories for each store.

6.5.2 Spatial Unit of Analysis

Our indicators of purchasing was generated at the level of Dissemination Area (DA), the smallest census administrative unit in Canada containing between 400 and 700 residents (145). The borders of DA align with boundaries of larger and commonly used administrative area for spatial analysis, such as census tract and municipality.

For disease risk mapping of T2D using our indicators (secondary Objective), the spatial unit of analysis was neighborhood that was formed by merging census tracts in a way to maximize the socio-demographic and economic attributes of residents (144). Therefore, the value of the indicators (area-level quantity of purchasing) was aggregated from 6,259 DAs into 193 neighborhoods by weighted averaging, using as weights the normalized area of overlap between each neighborhood and the spatially intersecting DAs. We selected the larger areal unit, neighborhood, for disease mapping, since this unit is likely to be an adequate representation of community for the delivery of interventions, but far smaller than the geography of inference (province or census metropolitan area) provided by the existing nutrition survey in Canada.

6.5.3 Statistical Analysis

The analysis followed three steps. In the first step, we predicted store-level sales of the four target food categories among out-of-sample food outlets in 2012 using a hierarchical Bayesian model. As the sales variables were highly non-normally distributed (right skewed), they were natural-log transformed. In the second step, we generated pairwise visit probabilities between each DA and food outlets given the distance and store size. Predicted (from out-of-sample stores) or observed (from sampled stores) store-level sales of a given food category were partitioned and allocated to DAs, with the amount proportional to the visit probability and population size in each DA. These are our small area indicators of food purchasing. In the third step, we evaluated the usefulness of these indicators in improving the fit of a disease mapping model to estimate area-level risk of prevalent T2D.

Sales prediction model

Prediction of sales for the out-of-sample stores was performed using a proper Conditional Autoregressive (CAR) model we have described previously (221); see Appendix 1. In short, we modelled the natural log-transformed average weekly sales of each food category as a function of store- and area-level predictors of sales. We did not jointly predict the sales of these food categories, as different types of stores carried different foods (e.g. the yogurt models did not include sales from pharmacies), thus there were different underlying data from each type of store with which to learn the sales.

Calculation of origin-destination probability and allocation of store sales into areas

The Huff gravity model uses travel cost and store size as the determinant of store selection to generate origin-destination probability (222,227). We assumed that shopping of residents in the CMA of Montreal is constrained to the stores within CMA (e.g. no spillover of shopping to stores outside of the CMA), which is partitioned into I = 6,259 DAs and contains 1,097 chain stores selling the target food categories (1,097 and 311 stores for sodas and yogurts, respectively), excluding chain convenience stores. For a given product category, the selection probability, π_{ij}, of residents in DA i choosing each alternative store is derived from a choice function as follows:

$$\pi_{ij} = \frac{A_j^{\gamma} d_{ij}^{-\lambda}}{\sum_{j=1}^{J} A_j^{\gamma} d_{ij}^{-\lambda}} \ ,$$

for $i = \{1, \cdots, I\}$. The number of potential stores to choose for residents in DA i is $\in \partial_i$, where ∂_i is the index set of stores located within 20-kilometer from the centroid of DA i based on road network distance. We considered it highly unlikely that residents would undertake grocery shopping beyond this distance. Each store therefore had a differing number of potential DAs to serve. Because the value of the attraction variable (A_j) varies across the four food categories of interest, the value of π_{ij} area also food category-specific.

The attraction of store, A_j, is traditionally measured by store size (e.g. square footage) devoted to the sale of a given food category with an attraction coefficient γ. As the measure of physical size was not available to us, we used the store-level assortment breadth of each food category (the number of distinct food items) in 2012. The measure of travel cost between the centroid of originating DA i and destination store j is denoted as d_{ij} with the corresponding decay coefficient λ, and was measured as the shortest street distance based on the 2013 Canada road network file (228).

Mean annual store-level sales of a given food category in store j were then allocated to DAs, and the portion of allocation was determined by the store visit probability and the number of residents in each area. $\sum_{j=1}^{J} \pi_{ij} = 1$ The number of residents moving from DA i to surrounding store j located within 20 kilometer of the area's centroid is $\pi_{ij} C_i$, where C_i is the total number of residents in DA i as measured by the 2011 Canadian Census. Observed or estimated (posterior mean) sales in terms of serving quantity for a store j is D_j. Using $\pi_{ij}C_i$ as normalized weights, the total quantity purchasing of a given category by residents in DA i was determined as $\sum_{j=1}^{J} D_j \pi_{ij} C_i$. Since this is the sum of partitioned product sales from J stores, areas containing no stores also received sales from stores in other areas. Note that although DA is a small spatial unit, multiple stores can belong to a single DA (e.g. shopping district), but these DAs may not receive a larger quantity of allocated sales from these stores than surrounding areas, if their number of residents is small.

It should be noted that the distribution of distance decay and attraction parameters are intended to be estimated from shopping-related mobility data (229,230). Unfortunately, such data are not available in our study due to the cost of administering or obtaining such a survey with adequate geographic accuracy allowing the identification of destination stores. For demonstrating the application of the Huff model to public health surveillance, we opted for a heuristic solution of adopting the commonly used ad-hoc value of 2.0 for the distance decay coefficient (thus power decay of 2.0), and value of 1.0 for the attraction coefficient as in the original study developing the Huff Model, thus making π_{ij} a deterministic quantity. As well, two DAs with the same population size and distance from a given store received the identical amount of partitioned sales, even if one area has a higher proportion of residents with a characteristic associated with higher soda sales (e.g. lower education).

Diabetes risk model

As described previously (221), we defined T2D cases as residents receiving at least one hospital diagnostic code (International Classification of Diseases [ICD] version 10) of T2D or at least two diagnoses of T2D in the physician claims database (ICD version 9) within a two-year period (221). The cases and non-cases are a 25% random sample of individuals in the CMA provided by the provincial public health insurance registry, the Regie de l'Assurance Maladie du Quebec, with exclusion criteria being under the age of 2 or having gestational diabetes. These prevalent cases were linked to respective neighborhoods in the CMA (n = 193) based on the 6-digit Canadian postal code.

The area-level aggregated T2D count $y_k = \{y_1, \ldots, y_n\}$ is assumed to follow a Poisson distribution with mean $e_k\theta_k$, where e_k is an offset term representing the age-adjusted expected number of cases calculated by an indirect standardization (Appendix 2). The area-specific relative risk adjusted for age, θ_k, represents the deviation of T2D risk from expected count and was modelled as:

$$log(\theta_k) = \alpha + G_k\gamma + X_k\beta + b_k \ ,$$

where α is a constant term, and G_k is a vector of area-level covariates associated with T2D with the corresponding coefficient vector γ. These covariates are education as the proportion of residents having post-secondary diploma or certificate among those over age 25 years, income as median family income, immigrant as the proportion of immigrants, and the availability of recreational facilities encouraging physical exercise as the number of facilities pe resident, as calculated from the 2011 Canadian National Household Survey. The location of recreational facilities were obtained from the CBP data and calculated as in our previous study (221). The vector of four elements representing the purchasing indicators and the corresponding coefficients in area k is denoted as X_k and β, respectively.

To account for spatial autocorrelation and Poisson overdispersion of the disease risk, we added an area-level random effect b_k, which is a re-parametrized version of the Besag-York-Mollie (BYM) CAR prior (152) as proposed by Riebler *et al.* (160). The BYM-2 prior is specified as:

$$b = \sigma_b\left(\sqrt{1-\phi}v + \sqrt{\phi}u\right),$$

where σ_b is a standard deviation of the sum two random effects: v as non-spatial random effect and u as an intrinsic CAR (ICAR) prior with a spatial structure (151), that is defined as a binary neighborhood weight matrix i.e. areas were considered as neighbors if they share common boundary. A mixing parameter ϕ describes relative contribution of the two random effects, with $\phi = 1$ indicating residual relative risk wholly explained by the spatial (ICAR) component. In the BYM-2 formulation, the variance of the two random effects were scaled to 1, such that the interpretation of the hyperparameter is independent from underlying neighborhood structure unlike the classic BYM CAR model, in addition to allowing the standardized comparison of hyperparameter for spatial and non-spatial random effect.

Sampling from the posterior distribution of the parameters was carried out by Hamiltonian Monte Carlo using the software package Stan. We examined convergence by investigating the trace plot of 3 independent chains. We discarded the first 5,000 samples as burn-in and retained the remaining 30,000 iterations as an inference sample to compute the posterior summary and model fit, which is the Watanabe-Akaike Information Criterion (WAIC). To investigate the potential consequence of excluding soda sales from convenience stores (yogurts are not sold in these stores) in the disease risk, we added a covariate representing the area-level number of convenience stores per resident in the T2D model as a sensitivity analysis. The prior probability of the parameters in the T2D model was $\gamma\ N(0, 5^2)$ for each regression coefficient, $\phi\ Uniform(0,1)$ for the mixing parameter, and a zero-mean half-normal distribution with the variance 5^2 for the common standard deviation of the random effect, σ_b.

6.6 Results

Table 6-1 provides the summary distribution of serving-standardized sales quantity of the target food categories. The variation of sales clearly reflects type of store: the stores selling yogurt did not include pharmacy chains (whereas pharmacies do sell sodas), resulting in a larger volume of minimum sales. Mean and median sales of soda and flavored yogurt were considerably larger than the low-calorie (diet) counterpart. Similar to sales, the availability (number of distinct items) of soda and flavored yogurt was larger than that of diet-soda and plain yogurt, respectively (Table 6-2). The summary distribution of store neighborhood SES from sampled stores (Table 6-3) is similar to all areas in the CMA of Montreal (Table 6-4), suggesting the representativeness of the sampled stores in terms of local SES attributes.

The estimated parameters of the sales models (Figure 6-1) indicates that chain-random effects had a considerably larger association with the sales of soda and diet-soda than the sales of yogurts. The fixed effects representing area-level characteristics, however, did not have notifiable association with the sales of sodas and yogurts, except that sales of plain yogurt was positively associated with store neighborhood education and negatively associated with store neighborhood income. In general, fitted and observed value of soda and diet-soda appeared to correlate well (Figure 6-2a and 6-2b). On the other hand, fitted sales of flavored and plain yogurt (Figure 6-2 c and 6-2d) seem to be underestimated at larger stores and overestimated at smaller stores, in addition to a larger uncertainty of the predictive distribution relative to that of sodas.

Figure 6-3 a)-d) shows the geographic distribution of the weekly purchasing of the target food categories per resident, aggregated from the smaller DA-level purchasing indicator as generated by the gravity model. While purchasing of both sodas and flavored yogurt appeared to be higher outside the island of Montreal, spatial distribution of plain yogurt was more homogeneous.

Table 6-5 shows the exponentiated mean and posterior credible interval of the area-level covariates, our area-level indicators of food purchasing, and the scaling and mixing parameters of the BYM2 model. The indicator for plain yogurt had a conclusive and negative association with the area-level relative risk of T2D (95% credible interval greater than 1), while that of flavored yogurt showed a negative association and nearly conclusive (lower bound of credible interval was 0.9999). On the other hand, the association of the indicator for both soda categories with the risk of T2D was inconclusive. Fit of the T2D model including our indicators was modestly superior to the model without the indicators (WAIC = 1769 vs. WAIC = 1772). Visual comparison of the fitted and observed count of T2D prevalence (Figure 6-4a and b) reveals an already good fit of the model without the indicators. The marginal variance of the random effect was slightly reduced upon addition of our indicators, and the decreased posterior mean of the mixing parameter,ϕ, indicates that our indicator in fact reduced spatial component of the variance. Regardless, much of the variance was attributed to non-spatial component.

The number of convenience stores per resident to the T2D model did not show conclusive association (posterior mean, 1.01 [95% credible interval: 0.98, 1.03]), and did not notably change the posterior mean nor the credible interval of our purchasing indicators.

6.7 Discussion

We demonstrated the application of the Huff retail gravity model to grocery POS data in order to address the current lack of small-area measurement for community diets, a critical public health indicator to allow planning, targeting and evaluating nutrition interventions. As the addition of the purchasing indicators improved fit of the area-level T2D model for disease mapping, we demonstrated the usefulness of our indicators in assisting small area assessment of disease burden associated with community dietary patterns

The posterior summaries of the sales models indicate that the chain-level factor has a strong association with the sales of soda and diet-soda, but less so with the sales of plain and flavored yogurt. It is possible that chain-level marketing strategies for yogurt are not strongly differentiated across supermarkets and/or purchasing behavior is independent of chain-level factors. It should also be noted that yogurt models did not include pharmacies whose chain random effects showed more prominent difference. The slightly biased nature of the predicted yogurt sales can be seen in the fitted and observed sales plot, which suggests the omission of important store or area-level attribute(s) for these products.

Visually similar spatial distribution of the indicator for soda and diet-soda may suggest the absence of regional preference for diet products. However, the spatial distribution of purchasing for plain yogurt was distinct from that of high-caloric (flavored) yogurt. Regardless of the spatial patterns, purchasing quantity of flavored yogurt was far greater than that of plain yogurt for all areas, a finding consistent with yogurt being ranked in the top 10 source of total sugar intake among Canadian children in 2015 (53). Interestingly, the center of the island showed a cluster of areas with a relatively low level of purchasing for soda, diet-soda and flavored yogurt, even though these areas have high density of food outlets.

Our purchasing indicator of plain yogurt was associated with decreased relative risk of T2D prevalence, while positive association was observed for the purchasing of flavored yogurt, although the latter association was not conclusive. Accumulating evidence suggests potential benefits of yogurt in preventing numerous NCDs including T2D (61), but the healthful effect is likely to be limited to plain yogurt that is free from artificially added sugar. Although there is no epidemiologic evidence directly linking flavored yogurt with chronic diseases, consumption of flavored yogurt is less preferred over plain yogurt, as the former can lead to excess sugar intake that is unequivocally linked to obesity and overweight, and a likely risk factor for T2D. Aligned with this, our finding implicates the monitoring of yogurt items (and likely other food products) separately in accordance nutrient composition relevant to health risk.

Measuring diets for small areas using population health surveys is not feasible routinely or on a large scale due to the expense of achieving adequate sample sizes. Large-scale epidemiologic cohorts to allow adequate precision at modestly small area exist (22), but manual and regular data collection is not sustainable for decades of nutrition monitoring. Although very large in sample size, two recent applications of individual shopper-level POS data to generate a small area indicator of diets were limited to data from a single major supermarket chain (232,233). Since consumer selection of chain is differential across population subgroups with a varying risk of chronic diseases (234,235), inferences in these studies have limited representativeness and potentially exclude groups with a higher body mass index who may shop primarily in discount superstore chains (236). To our knowledge, ours is the first study to demonstrate an approach to overcoming the sparseness of survey sampling and non-representativeness of the shopper-level POS data, as we generated the indicators using store-level sales data from multiple chains, which were not limited to supermarkets. Additionally, the two studies above and the majority of research in public health nutrition has not adequately addressed spatial autocorrelation of health outcomes (98). Our sales and T2D risk model provided appropriate analytical approaches to account for latent spatial effects due to unmeasured variables with spatial structure, while the latter model overcomes the shortcomings of the classic disease mapping (BYM) model by providing means for standardized interpretation of the hyperpriors for random effects.

A major limitation of our study is the deterministic nature of the indicators, as we did not calibrate the distance decay and attraction parameters of our retail gravity model using mobility data, therefore treating the value of these parameters known *a priori* without uncertainty. Because representative origin-destination survey recording consumer selection of specific store was unavailable, we were limited to using commonly applied values from previous research. Nevertheless, we illustrated an application of retail gravity model to advance measurement capacity of nutrition surveillance.

Future research should aim to generate indicators for a comprehensive set of food products that represents healthiness of diets, including purchasing of fruits and vegetables, whole grains, nuts and seeds, and milks as well as other highly processed food related to diets high in sugar, fat and sodium. Although the law of retail gravitation suggests that travel distance and store size as the key factor of store selection, there are additional attributes that maybe associated with the probability of store visit, such as availability of parking, pricing (discount chain or not) (237,238) and being in shopping mall that should be added to the gravity model. Additional work includes linking individual food items to nutrition composition database to measure the purchased quantity of macronutrient (e.g. sugar) as a better predictor of disease risk than the quantity of servings. This is particularly important for flavored yogurt, as it constitutes a highly varying amount of added free sugars across items and therefore should not be simply dichotomized as flavored or plain.

Effective public health strategies to improve population diets are urgently needed to reduce future NCD burdens and nutrition disparity. There is a growing priority in incorporating socio-cultural and built environmental determinants of heath and in delivering nutrition interventions tailored to community needs. Geographic public health surveillance plays a primary role in guiding public health actions in this context and benefits from our methodology applied to POS data.

Table 6-1. Characteristics of Store-level Transactions by Food Category, Sampled Stores in the CMA of Montreal, 2012

	Min	Mean	Median (IQR)	Max
Soda	8.4	18,510.3	13,071.6 (1,322.7-26,957)	108,771.0
Diet soda	2.6	6,939.4	5,214.5 (664-11,588.9)	29,466.2
Flavored yogurt	1679.4	6,809.6	6,454.6 (4,479.4-8,711)	17,361.3
Plain yogurt	81.3	698.9	588.2 (333.4-907.3)	2,556.3

Abbreviations: IQR, interquartile range; Max, maximum; Min, minimum

Note: These quantities are average of weekly beverage transaction in serving
Sodas were sold in 1097 stores consisting of supermarket, supercenter, and pharmacy.
Yogurts were sold in 311 stores consisting of supermarket.

Table 6-2. Business Attraction of Retailers as Represented by Average of Weekly Number of Distinct Items Appeared in Transaction by Food Category, Sampled Stores in the CMA of Montreal, 2012

	Min	Mean	Median (IQR)	Max
Soda	1.3	46.6	58.4 (18.7-69.6)	92.3
Diet soda	2	41.6	50 (28-56)	80
Flavored yogurt	45	129.2	120.3 (111.6-159.5)	189.7
Plain yogurt	5.7	16.6	16.1 (13.2-20.5)	28.8

Abbreviations: IQR, interquartile range; Max, maximum; Min, minimum

Note: Sodas were sold in 1097 stores consisting of supermarket, supercenter, and pharmacy.
Yogurts were sold in 311 stores consisting of supermarket.

Table 6-3. Characteristics of Neighborhoods on Which Sampled Chain Stores Were Located (83 areas) in Census Metropolitan Area of Montreal, 2011 Canadian National Household Survey

Neighborhood characteristics	Min	Mean	Median (IQR)	Max
Education [a]	29.6	68.0	67.6 (61.5-74.8)	90.3
Median family income (in 10,000 Canadian dollars)	2.0	7.4	7.0 (5.8-8.7)	16.7
% of immigrant	1.6	21.6	20.4 (8.1-32.6)	63.1
% of residents under 18 years old	2.4	19.7	20.5 (17.2-23.3)	32.9
% of employed residents among labor force	18.5	59.9	59.0 (54.1-66.9)	76.8
Average family size	1.1	2.9	3.0 (2.9-3.1)	3.5
Population density (residents per square kilometer)	109.2	3,811.8	2,298.6 (1,074.1-	19,606.1

Abbreviations: IQR, interquartile range; Max, maximum; Min, minimum

[a] % of residents with post-graduate diploma or certificate among age greater than 25

Table 6-4. Characteristics of Neighborhoods (193 areas) in Census Metropolitan Area of Montreal, 2011 Canadian National Household Survey

Neighborhood characteristics	Min	Mean	Median (IQR)	Max
Count of T2D	43.0	346.1	319.0 (211.0 -	1617.0
Prevalence of T2D (%)	3.1	7.0	6.8 (5.8-8.2)	10.7
Education [a]	29.6	67.1	66.3 (60.5-74.8)	90.3
Median family income (in 10,000 Canadian dollars)	2.0	7.2	6.9 (5.5-8.5)	16.7
% of immigrant	1.6	22.6	20.9 (9.5-32.7)	63.6
% of residents under 18 years old	2.4	19.9	20.9 (17.5-23.3)	32.9
% of employed residents among labor	18.5	59.0	59.0 (54.1-66.3)	76.8
Average family size	1.1	2.9	3 .0 (2.8-3.1.0)	3.5
Population density (residents per square kilometer)	54.7	4,023.5	3,348 (1,093.5-5,744.8)	19,606.1

Abbreviations: IQR, interquartile range; Max, maximum; Min, minimum; T2D, Type 2 Diabetes Mellitus

[a] % of residents with post-graduate diploma or certificate among age greater than 25

Table 6-5. Posterior Mean and 95% Credible Interval of Exponentiated Coefficients in Diabetes Risk Model, Census Metropolitan Area of Montreal, 2012.

	No purchasing indicators		With purchasing indicators	
Parameter	Mean	95%CI	Mean	95%CI
Intercept	1.10	(1.03 , 1.17)	1.15	(1.07 , 1.23)
Area-level attributes (scaled)				
Education	0.92	(0.89 , 0.95)	0.95	(0.92 , 0.99)
Immigrant	1.06	(1.03 , 1.10)	1.09	(1.05 , 1.13)
Income	0.93	(0.90 , 0.96)	0.90	(0.87 , 0.93)
Recreation	0.98	(0.96 , 1.00)	0.97	(0.95 , 0.99)
Sales (scaled)				
Soda			1.01	(0.98 , 1.04)
Diet soda			0.99	(0.93 , 1.06)
Yogurt(plain)			0.94	(0.91 , 0.97)
Yogurt(sugar)			1.07	(0.999998 ,1.14)
BYM parameters				
σ_b	1.13	(1.11 , 1.15)	1.12	(1.10 , 1.14)
ϕ	0.40	(0.10 , 0.73)	0.33	(0.07 , 0.66)

Abbreviations: 95% CI, 95% credible interval; DIC, deviance information criterion; Mean, posterior mean; Recreation, Recreational facility per resident.

[a] Covariates were mean centered and scaled to one standard deviations, and the regression coefficients were exponentiated. The value of coefficients represents an area-level relative risk of T2D, which is the ratio of the risk at one unit increase of the covariates to the risk at mean value of the covariates.

Note: The BYM-2 parameters were not exponentiated.

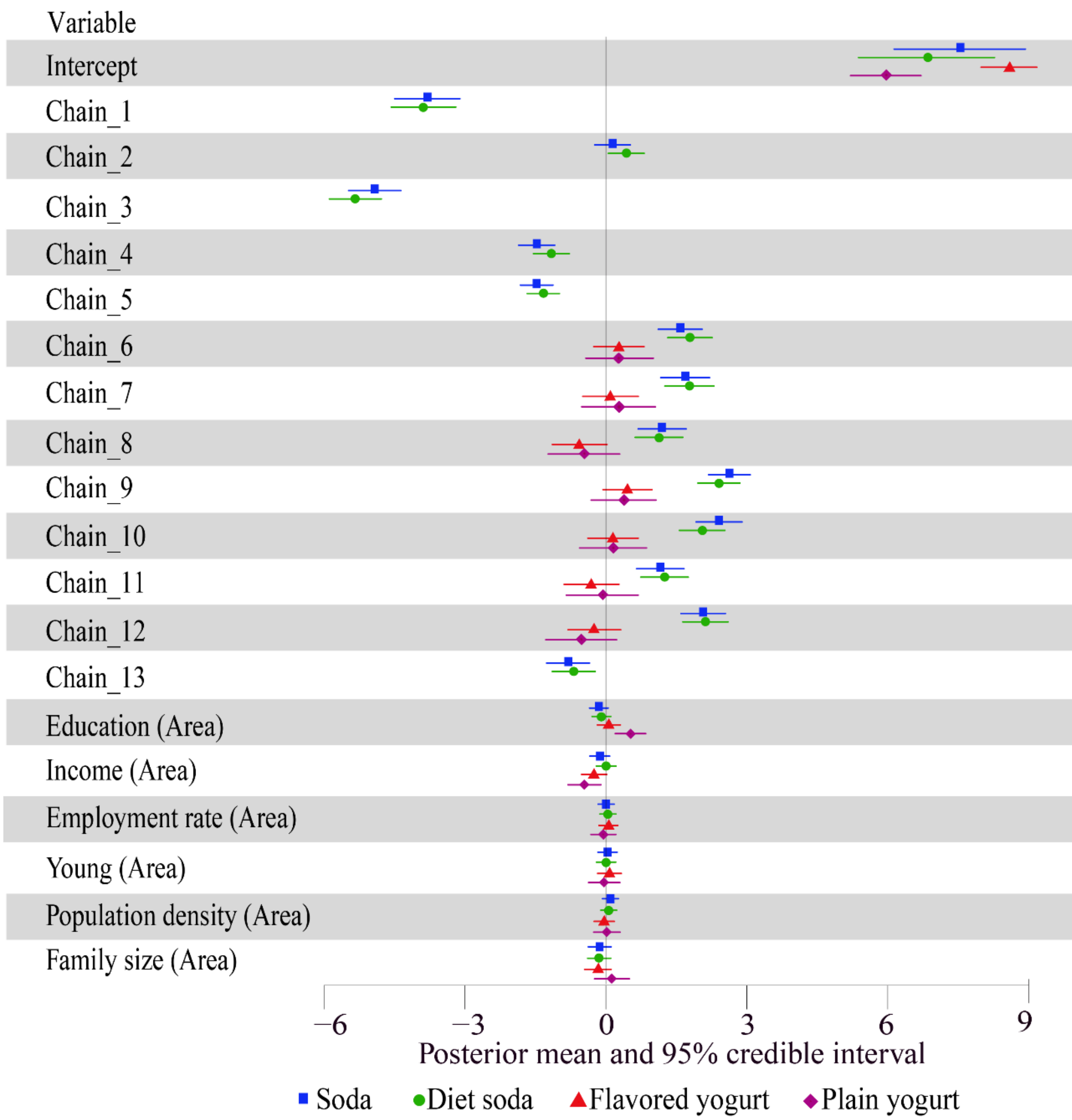


Figure 6-1. Posterior Summary of Sales Model for Each Food Category.

Models were ran separately. Chain 1 to 5 are Pharmacy, Chain 6 to 11 are Supermarket¬, and Chain 12 and 13 are Supercenter

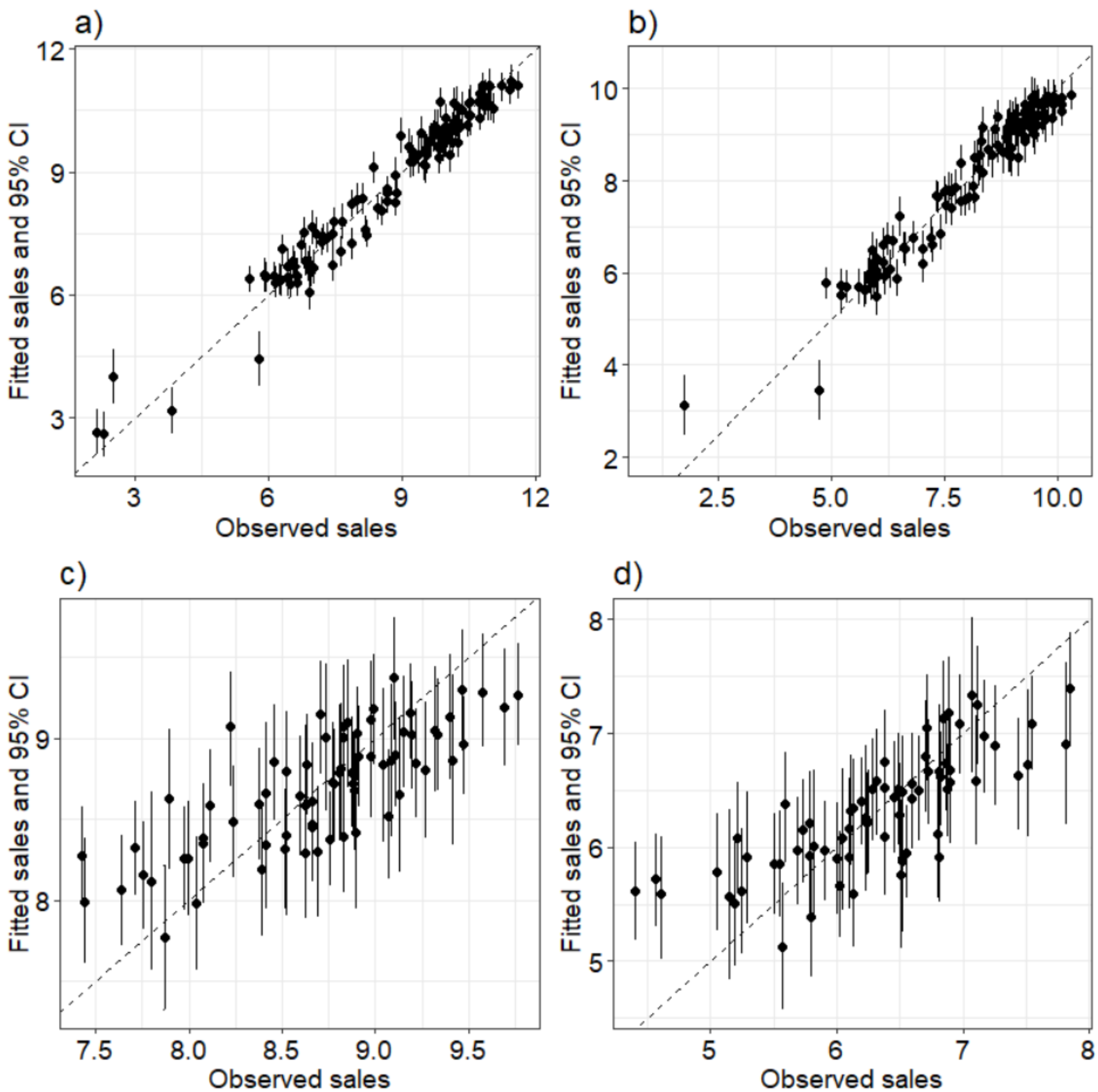


Figure 6-2 A) –D). Fitted and Observed Log Sales of a) Soda, b) Diet-soda, c) Flavored Yogurt and d) Plain Yogurt in the Sales Model.

Note that axis scales vary by product category

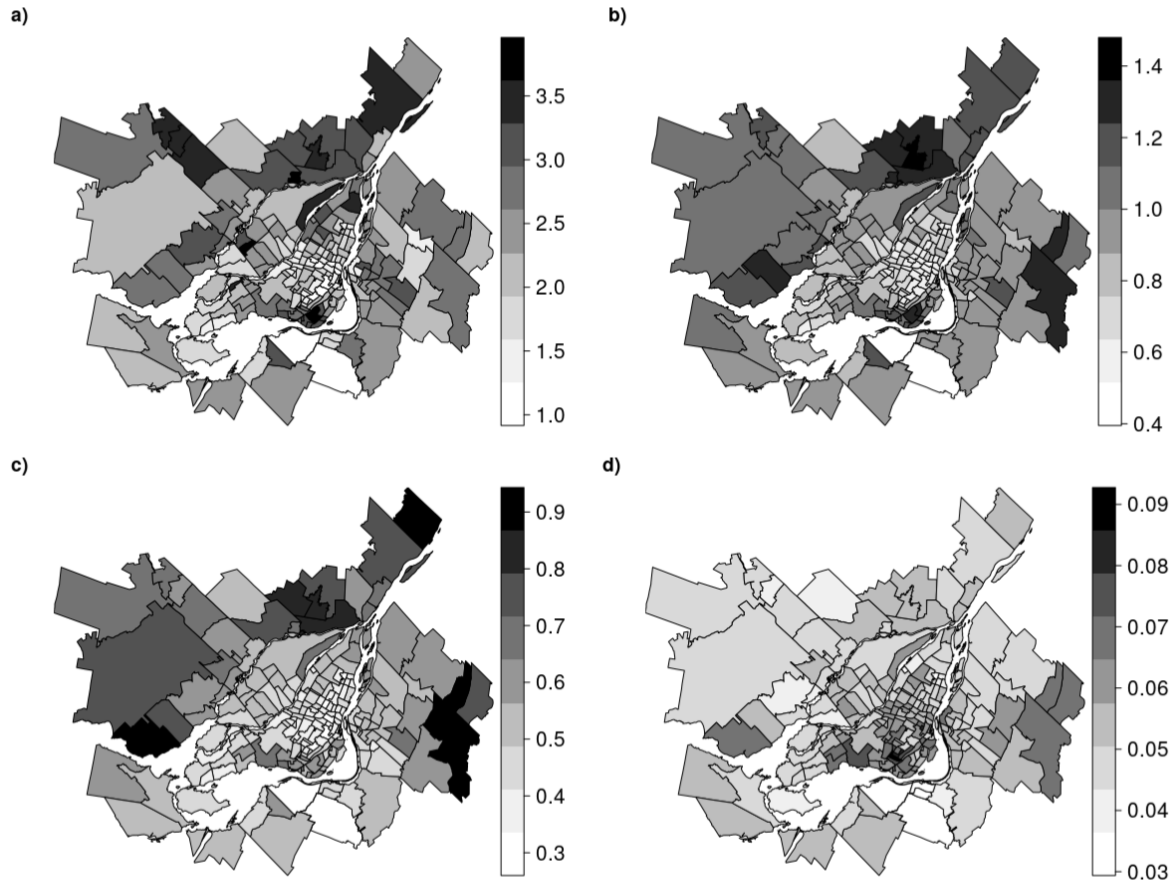


Figure 6-3 A)-D). Spatial Distribution of Decomposed Purchasing Indicator for a) Soda, b) Diet-soda, c) Flavored Yogurt, and d) Plain Yogurt in Census Metropolitan Area of Montreal, 2012

Note: The scales are not standardized. The indicators were calculated based on the re-distribution of predicted and observed sales by population count and store visit probability, fixing the sale of each stores.

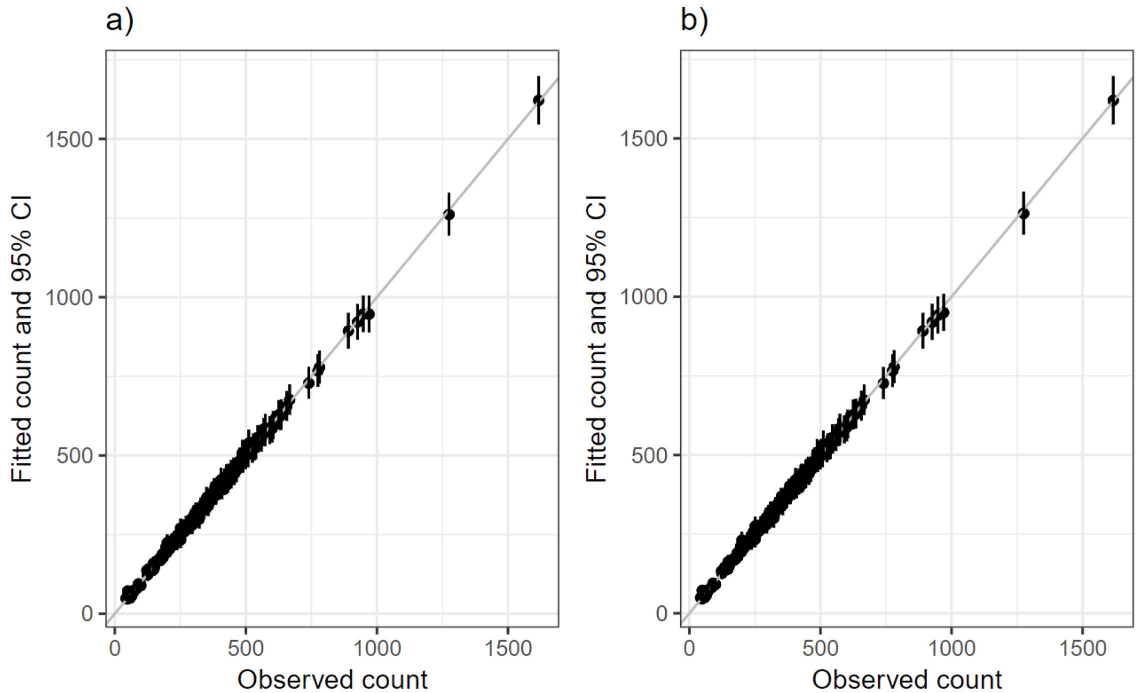


Figure 6-4 A) and B). Fitted and Observed Count of Area-level Type 2 Diabetes Mellitus for Model A) Without and B) With Purchasing Indicators

6.8 Supplementary Materials for Manuscript 3

Appendix 1. Specification of Sales Model

We first calculated store-level average weekly sales of each food category in 2012, which was natural log-transformed to follow a normal distribution. For a given food category, we assumed that the average weekly sales at store j located in area i (D_{ij}) were generated by a hierarchical normal distribution with mean μ_{ij} and variance σ^2. We decomposed the mean as the sum of store and area-level characteristics by letting,

$$\mu_{ij} = \beta_0 + \phi_{c[j]} + Z_i$$

$$Z_i = A_i\varphi + S_i \ ,$$

where the component $\phi_{c[j]}$ is a random effect capturing the association of sales with store chain c to which store j belongs. Area-level attributes in the model are captured by Z_i, with the component A_i representing a vector of neighborhood attributes in an area containing the store, φ is a vector of the corresponding coefficients , and S_i is a spatially structured random effect accounting for the remaining spatial variation in sales after accounting for the store and area-level characteristics.

Specifically, A_i includes area-level median family income, education as the proportion of individuals over age of 25 with post-graduate certificate or diploma, population density as the number of residents per square kilometer, proportion of residents under 18-years old, family size as the mean number of family members, and employment rate as the proportion of those in the labour force employed in full- or part-time work, obtained from the 2011 National Household Survey (1). All the covariates were standardized to have a zero mean and standard deviation of one.

The spatial random effect $S = (S_1 \cdots, S_n)^T$ captures the spatial dependency of sales across areas using a proper Conditional Autoregressive (CAR) prior (2), with a binary spatial weight matrix where areas that share boundaries are defined as neighbors.

Prior distribution selected for regression parameters, φ and β_0 was $Normal(0,100)$, and the prior distribution for precision parameters, σ_{ϕ}^{-2}, σ^{-2}, and the variance parameter of the proper CAR prior, τ^{-2}, was $Gamma(0.5,0.01)$. Sampling from the posterior distribution of the parameter vector was done using Markov chain Monte Carlo through OpenBUGS software (3). We let the MCMC run for 60,000 iterations, considered the first 10,000 as burn-in and stored every 20th iteration to avoid autocorrelation among the sampled values. Monte Carlo Standard Error was less than 5% of sample deviation. Trace plots of three MCMC chains starting with very different initial values were visually inspected to investigate convergence.

Appendix 2. Calculation of Age-Adjusted Expected Count of Diabetes.

Age-adjusted expected count for area *i* was calculated by indirect standardization;

$$e_i = \sum_{k=1}^{8} N_{ik}\, p_k \ ,$$

where N_{ik} is the population size greater or equal to age two in area *i* at age stratum *k*.

Age strata k were defined as [2-20) [20-30) [30-40) [40-50) [50-60) [60-70) [70-80), and>=80 years old. Background (reference) disease prevalence in age stratum k in the study region (CMA of Montreal) was denoted as p_k and calculated as the number of people having diabetes status in 2012 in stratum k divided by the corresponding age stratum-specific total population with and without diabetes diagnosis.

References for Supplementary Materials

1. Government of Canada SC. 2011 National Household Survey Profile - Census metropolitan area/Census agglomeration. May, 8, 2013 accessed 2016 May 16. https://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMA&Code1=462&Data=Count&SearchText=462&SearchType=Begins&SearchPR=01&A1=All&B1=All&Custom=&TABID=3

2. Banerjee S, Carlin B, Alan G. Hierarchical Modeling and Analysis for Spatial Data, Second Edition. 2nd ed. Boca Raton, FL: CRC Press; 2014.

3. Lunn D, Spiegelhalter D, Thomas A, et al. The BUGS project: Evolution, critique and future directions. Stat Med. 2009;28(25):3049–67.

Chapter 7: Discussion and Conclusions

7.1 Summary of Main Findings

Manuscript 1 showed that the extent of the interaction between store neighbourhood education and price discounting varied markedly across store types. The heterogeneity of community behavior implies that monitoring or investigating point-of purchase patterns only in a single store type (typically supermarkets in food environment research) can lead to erroneous conclusions, if capturing population-representative behaviors of food selection is the research aim. For this reason, increasing prevalence and market share of non-traditional retail format (e.g. pharmacy and supercenter) must be recognized in future research.

The negative and conclusive effect of the interaction term found in the pharmacy subsample indicates the strongest effect of discounting on soda purchasing in communities with the lowest tertile of education. The results thus suggest the possible contribution of price discounting to the mechanism creating disproportionately worse diets among deprived communities.

Although the interaction was not conclusive in supermarkets and supercenters, price discounting was found to increase soda sales in all store types, regardless of the levels of store neighbourhood education. Therefore, price discounting maybe an important consumer environmental risk factor of unhealthy food purchasing, irrespective of socio-economic status.

Manuscript 2 proved the feasibility of our analytical approach in generating a small area indicator of (latent) soda demand from POS data. The addition of the proposed indicator to a T2D risk model improved model fit, suggesting that the information extracted from POS data has a potential to help improve the assessment of local disease burden. Importantly, the spatial smoothing property of a CAR prior solved the main challenge of using POS data for small area estimation of diets: absence of food outlets in some areas.

The proposed indicator also showed a superior model fit over the model containing mRFEI, a commonly used indicator of community food environment. In the absence of a small-area measure of diet, researchers in geographic disease surveillance often used built environmental measures of physical accessibility to stores, such as mRFEI, to predict areal disease risk of NCDs. However, consistent association between spatial accessibility to stores and diets is lacking to date, and therefore these store location-based food environmental measure cannot be a surrogate of local diets. In addition, the proposed indicator of food demand can be estimated in areas lacking stores, while mRFEI cannot be calculated in areas without stores.

The results of the soda sales model suggest that the sales of soda are mainly driven by chain-specific characteristics as represented by the strong effects of the chain random effects. In contrast, store neighbourhood socio-demographic and economic indicator that had inconclusive association with sales, except education.

Manuscript 3 proved the utility of an alternative approach generating small area indicator of diets: the decomposition of store-level sales into area-level purchasing using origin-destination probabilities generated by the Huff gravity model. As in Manuscript 2, the resulting indicators improved model fit of the T2D risk model for disease mapping.

The spatial distribution of the indicator for plain (non-sugar) and flavored (sugary) yogurt was considerably different, and their association with T2D risk was opposite in direction. The observed positive association of plain yogurt purchasing with T2D risk is concordant with the recent public health concern about the nutritional risk of flavored yogurt in promoting diets high in free sugar. The finding indicates that defining and monitoring food categories by nutritional composition (free sugar in the case of yogurt) maybe crucial. On the other hand, the small area indicators of soda and diet-soda purchasing showed a nearly identical spatial distribution to each other, although their association with T2D risk was highly inconclusive. Soda and diet-soda were purchased far greater than yogurts (with plain yogurt being least popular) in terms of quantity for all areas.

7.2 Strengths

The primary strength of this thesis is the development of new methods for harnessing relevant features in POS data: purchasing quantity, time (week), location, and promotional activities associated with many food items.

Although research investigating the association of price discounting and purchase responses has begun to appear in recent years, these studies have tended to aggregate over temporal variation of pricing and sales, and are therefore cross-sectional in design (17,85,239). In addition, these studies analyzed pricing records generated by one store type (i.e. supermarket) or a single supermarket chain. In light of heterogeneous purchasing patterns across store types shown in Manuscript 1, my thesis captured more population-representative community responses to price discounting.

While public health use of electronic grocery transaction data to capture consumer or community behaviors is increasing (240), spatial analysis of these data has not been performed to date. Geographic patterning of food purchasing therefore remains unknown. Manuscript 2 provided a principled application of spatial statistical analysis, while Manuscript 3 illustrated the usefulness of retail gravity analysis to generate area-level measures of purchasing from POS data. Both approaches generated a purchasing measure for areas lacking stores. The two manuscripts not only provide plausible methods to generate much needed community indicators of diets, but they also enrich the application of spatial analysis in food environment research, where most analyses fails to acknowledge spatial dependency of measurements.

7.3 Limitations

The POS data in this thesis excluded sales generated by independent stores, warehouse clubs, fastfood restaurants, and stores not equipped with product scanners. Since the excluded store types offer price discounting far less frequently and had a smaller market share (proportion of sales) than the included store types, the exclusion may not be particularly critical in characterizing the role of price discounting in Manuscript 1. For Manuscript 2 and 3 however, it is possible that there are areas whose store composition is dominated by the excluded store types, and not including the sales from these stores may have biased (underestimated) the quantity of soda and yogurt purchasing in these areas.

Another potential source of bias in the latter two manuscripts is the exclusion of convenience stores, as these stores could not be assigned to the neighbourhood to which they belonged due to the unavailability of exact store locations.

As well, bias in the quantity of area-level purchasing in Manuscript 3 may have arose from inaccurate prediction of sales in out-of-sample stores, which is evident in the suboptimal agreement between the fitted and observed sales of yogurt. The biased prediction of sales may be due to unmeasured store and area-level factors associated with yogurt sales. In addition, the origin-destination probabilities generated by the Huff model in Manuscript 3 may not accurately reflect store visits of the Montreal population, as the distance decay and attraction parameters were not estimated by travel survey.

7.4 Future Work

Comprehensive assessment of dietary patterns requires the inclusion of multiple food categories, in addition to the two food groups targeted in my thesis (soda and yogurt). The methodologies in this thesis can be readily extended to other food categories including whole grains, fruits and vegetables, low-fat milk, and numerous ultra-processed food products. This could be accomplished through joint modelling of multiple time-series using the approach for Manuscript 1, as sales of food categories tend to correlate each other over time. As for Manuscript 2 and 3, indicators representing purchasing of multiple food categories would allow a better characterization of community diets and potentially lead to an improved fit of disease mapping models. In addition, an analysis considering multiple food categories would allow clustering of areas by dietary patterns, which may reveal within-cluster shared characteristics that could be a target of interventions.

Because the influence of price discounting on sales is mainly short-term in duration for soda, Manuscript 1 investigated the association of discounting and sales in the same week. However, sales of other food categories may experience a long-term increase even after the end of price discounting or a rapid reduction below the baseline (pre-discounting) sales due to stockpiling during the discounting period i.e. borrowed sales from the future (241–243). As well, it should be noted that price discounting of one food category may affect the short and long-term sales of other food categories due to across-category competition. A better public health insight in the role of price discounting would therefore require the estimation of time-varying effects using econometric time-series analysis, such as dynamic linear models.

Manuscript 2 and 3 are cross-sectional in design, focusing on the spatial analytical aspect of POS data. An immediate and necessary extension is the incorporation of temporal variation into the proposed community indicators of diets, for example at the scale of month, quarter or year. Such measures will allow health agencies, communities and researchers to assess the trajectories of neighbourhood diets in response to (or as a control group of) community and national interventions and socio-economically significant events, for example economic recessions. This is a notable improvement over the temporal measurement frequency of the Canadian national nutrition survey, which is repeated every 10 years.

7.5 Public Health Implications

The three Manuscripts provide novel approaches and evidence to support dietary surveillance and interventions to improve population nutrition. As one of the first studies to reveal the effect of price discounting on nutritional disparities, Manuscript 1 should stimulate the development of further observational and quasi-experimental research to inform policy formulation. Pharmacies were previously discussed as an emerging point of community-rooted health interventions (196,244). Manuscript 1 however, suggest a potential pathway through which pharmacies may have a negative effect on population nutrition, specifically as a provider of unhealthy food in a retail setting where discounting and low levels of education interact to increase purchasing. Further research is required to elucidate the mechanism through which discounting acts on SES-specific purchasing of food items in pharmacies.

Manuscript 2 and 3 advance the science of epidemiology and community intervention and the practice of surveillance by developing feasible, novel approaches to deriving population inference about nutrition from routinely generated commercial data. Measurement of food selection is a critical indicator to assess the effectives of nutrition interventions, which are ideally evaluated using primary and secondary data (245). Ad-hoc, primary collection of survey data is costly and unrealistic for many local interventions. However, in my thesis, I have demonstrated how secondary data, with a focus on POS data, can be used to measure and monitor community-level diet. Although further methodological refinements are needed, community organizations and researchers can use the approaches I developed to generate and routinely update community measures of diet. Public health agencies can also use the approaches I developed to conduct geographic surveillance of nutrition and NCDs, allowing generation of an atlas of neighbourhood diets, information necessary to characterize communities with the highest needs.

7.6 Conclusions

Nutrition-related NCDs are a leading cause of premature deaths and they disproportionately affect socio-economically disadvantaged individuals, communities, and countries. Preventive population health strategies targeting socio-economic and environmental determinants of diets are needed to reduce the burden and disparities of NCDs. Because national nutrition policies do not provide uniform benefits over population subgroups, minimization of nutrition disparity requires community-specific interventions to address the contexts that affect healthy eating. These interventions require regularly collected measures of nutrition-related behaviors and measures of environmental characteristics that shape food selections.

Price discounting is a powerful retail environmental driver of unhealthy food purchasing, but its impact on nutrition disparity has not been captured to date. Furthermore, geographic indicators of diet to assess community responses to nutrition interventions are lacking. My thesis develops and applies innovative solutions for these measurement gaps stemming from the limitations of traditional data sources in public health nutrition. Using weekly records of product marketing and sales available in POS data, Manuscript 1 evaluated SES-specific association of price discounting and soda sales. Using store location information in POS data, Manuscript 2 and 3 demonstrated novel approaches to small area estimation of dietary patterns.

As with the secondary use of other electronic data sources, effective use of POS data to answer existing public health inquiries requires novel research methods to extract relevant information. The three manuscripts have advanced measurement capacity in public health nutrition and demonstrated the transformative potential of POS data in help achieving precision public health, where tailored interventions that meet community needs are delivered in a timely manner using the best available evidence and local data.

Chapter 8: References

1. GBD 2016 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017 Sep 16;390(10100):1345–422.

2. Forouzanfar MH, Afshin A, Alexander LT, Anderson HR, Bhutta ZA, Biryukov S, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1659–1724.

3. World Health Organization. Global Action Plan for the Prevention and Control of NCDs 2013-2020 [Internet]. WHO. [cited 2019 Oct 7]. Available from: http://www.who.int/nmh/events/ncd_action_plan/en/

4. Satia JA. Diet-related disparities: understanding the problem and accelerating solutions. J Am Diet Assoc. 2009 Apr;109(4):610–5.

5. Blüher M. Obesity: global epidemiology and pathogenesis. Nat Rev Endocrinol. 2019 May;15(5):288.

6. Zimmerman FJ. Using Marketing Muscle to Sell Fat: The Rise of Obesity in the Modern Economy. Annu Rev Public Health. 2011;32(1):285–306.

7. Glanz K, Sallis JF, Saelens BE, Frank LD. Healthy nutrition environments: concepts and measures. Am J Health Promot AJHP. 2005 Jun;19(5):330–3, ii.

8. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. The Lancet. 2011 Sep 2;378(9793):804–14.

9. Diet, nutrition and the prevention of chronic diseases. Report of the joint WHO/FAO expert consultation [Internet]. WHO. [cited 2014 Dec 12]. Available from: http://www.who.int/dietphysicalactivity/publications/trs916/download/en/

10. Lytle LA. Measuring the food environment: state of the science. Am J Prev Med. 2009 Apr;36(4 Suppl):S134-144.

11. Richard L, Gauvin L, Raine K. Ecological Models Revisited: Their Uses and Evolution in Health Promotion Over Two Decades. Annu Rev Public Health. 2011;32(1):307–26.

12. Diez Roux AV, Mair C. Neighborhoods and health. Ann N Y Acad Sci. 2010 Feb 1;1186(1):125–45.

13. Brownson RC, Bright FS. Chronic disease control in public health practice: looking back and moving forward. Public Health Rep. 2004;119(3):230–8.

14. Government of Canada, Health Canada. Food and Nutrition Surveillance in Canada: An Environmental Scan [Internet]. 2005 Jul [cited 2014 Oct 5]. Available from: http://www.hc-sc.gc.ca/fn-an/surveill/environmental_scan-eng.php

15. Powell LM, Kumanyika SK, Isgor Z, Rimkus L, Zenk SN, Chaloupka FJ. Price promotions for food and beverage products in a nationwide sample of food stores. Prev Med. 2016 May;86:106–13.

16. Chandon P, Wansink B. Does food marketing need to make us fat? A review and solutions. Nutr Rev. 2012 Oct 1;70(10):571–93.

17. Nakamura R, Suhrcke M, Jebb SA, Pechey R, Almiron-Roig E, Marteau TM. Price promotions on healthier compared with less healthy foods: a hierarchical regression analysis of the impact on sales and social patterning of responses to promotions in Great Britain. Am J Clin Nutr. 2015 Mar 1;808–816.

18. Backholer K, Sacks G, Cameron AJ. Food and Beverage Price Promotions: an Untapped Policy Target for Improving Population Diets and Health. Curr Nutr Rep. 2019 Sep;8(3):250–255.

19. Waller LA. Mapping in Public Health. In: Brunn SD, Dodge M, editors. Mapping Across Academia. Dordrecht: Springer Netherlands; 2017. p. 169–81.

20. Li W, Kelsey JL, Zhang Z, Lemon SC, Mezgebu S, Boddie-Willis C, et al. Small-area estimation and prioritizing communities for obesity control in Massachusetts. Am J Public Health. 2009 Mar;99(3):511–519.

21. Estimated County-Level Prevalence of Diabetes and Obesity—United States, 2007. JAMA. 2010 Mar 10;303(10):933–5.

22. Dekker LH, Rijnks RH, Strijker D, Navis GJ. A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands – the Lifelines cohort study. Int J Behav Nutr Phys Act. 2017 Dec 7;14(166).

23. Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel JF, Clark A, et al. Disease mapping models: an empirical evaluation. Disease Mapping Collaborative Group. Stat Med. 2000 Sep 15;19(17–18):2217–41.

24. Government of Canada, Statistics Canada. Canadian Community Health Survey - Nutrition (CCHS) [Internet]. 2014 [cited 2018 Feb 18]. Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5049

25. James WHM, Lomax N, Birkin M. Local level estimates of food, drink and tobacco expenditure for Great Britain. Sci Data. 2019 May 13;6(1):1–14.

26. Nielsen. Retail Measurement: in-House Retail Experts [Internet]. [cited 2018 Apr 12]. Available from: http://www.nielsen.com/id/en/solutions/measurement/retail-measurements

27. Wedel M, Kannan P k. Marketing Analytics for Data-Rich Environments. J Mark. 2016 May 23;80(6):97–121.

28. World Health Organization. Non communicable diseases [Internet]. [cited 2019 Sep 13]. Available from: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

29. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017 Sep 16;390(10100):1151–210.

30. Hu FB. Globalization of Diabetes: The role of diet, lifestyle, and genes. Diabetes Care. 2011 Jun 1;34(6):1249–57.

31. Albright AL, Gregg EW. Preventing Type 2 Diabetes in Communities Across the U.S.: The National Diabetes Prevention Program. Am J Prev Med. 2013 Apr 1;44(4, Supplement 4):S346–51.

32. Government of Canada, Public Health Agency of Canada. Diabetes in Canada: Facts and figures from a public health perspective [Internet]. aem. 2011 [cited 2019 Sep 13]. Available from: www.canada.ca/en/public-health/services/chronic-diseases/reports-publications/diabetes/diabete

33. Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, Brauer M, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015 Dec 11;386(10010):2287–323.

34. Lieffers JRL, Ekwaru JP, Ohinmaa A, Veugelers PJ. The economic burden of not meeting food recommendations in Canada: The cost of doing nothing. PLOS ONE. 2018 Apr 27;13(4):e0196333.

35. Davis KA, Esslinger K, Elvidge Munene L-A, St-Pierre S. International approaches to developing healthy eating patterns for national dietary guidelines. Nutr Rev. 2019 Jun;77(6):388–403.

36. World Health Organization. Healthy diet [Internet]. [cited 2019 Sep 13]. Available from: https://www.who.int/news-room/fact-sheets/detail/healthy-diet

37. Moubarac J-C. Ultra-processed foods in Canada: Consumption, impact on diet quality and policy implications [Internet]. 2017 [cited 2019 Jul 20]. Available from: https://crdcn.org/ultra-processed-foods-canada-consumption-impact-diet-quality-and-policy-implications

38. Prospective Studies Collaboration, Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. Lancet. 2009 Mar 28;373(9669):1083–96.

39. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. Lancet. 2014 Aug 30;384(9945):755–65.

40. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017 [Internet]. p. 20. Available from: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf

41. Wijga AH, Scholtens S, Bemelmans WJ, de Jongste JC, Kerkhof M, Schipper M, et al. Comorbidities of obesity in school children: a cross-sectional study in the PIAMA birth cohort. BMC Public Health. 2010 Apr 9;10(1):184.

42. Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet. 2017 Dec 16;390(10113):2627–42.

43. Statistics Canada. Body composition of adults, 2012 to 2013 [Internet]. 2014 [cited 2019 Sep 13]. Available from: https://www150.statcan.gc.ca/n1/pub/82-625-x/2014001/article/14104-eng.htm

44. Statistics Canada. Body mass index of children and youth, 2012 to 2013 [Internet]. 2014 [cited 2019 Sep 13]. Available from: https://www150.statcan.gc.ca/n1/pub/82-625-x/2014001/article/14105-eng.htm

45. Rice K, Te Hiwi B, Zwarenstein M, Lavallee B, Barre DE, Harris SB. Best Practices for the Prevention and Management of Diabetes and Obesity-Related Chronic Disease among Indigenous Peoples in Canada: A Review. Can J Diabetes. 2016 Jun 1;40(3):216–25.

46. Hu FB. Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. Obes Rev. 2013;14(8):606–619.

47. Statistics Canada. Change in total sugars consumption among Canadian children and adults [Internet]. 2019 [cited 2019 Jul 21]. Available from: https://www150.statcan.gc.ca/n1/pub/82-003-x/2019001/article/00002-eng.htm

48. World Health Organization. Sugars intake for adults and children [Internet]. WHO. 2015 [cited 2015 Dec 28]. Available from: http://www.who.int/nutrition/publications/guidelines/sugars_intake/en/

49. Johnson RK, Appel LJ, Brands M, Howard BV, Lefevre M, Lustig RH, et al. Dietary sugars intake and cardiovascular health: a scientific statement from the American Heart Association. Circulation. 2009 Sep 15;120(11):1011–20.

50. Avena NM, Rada P, Hoebel BG. Evidence for sugar addiction: Behavioral and neurochemical effects of intermittent, excessive sugar intake. Neurosci Biobehav Rev. 2008;32(1):20–39.

51. Marriott BP, Olsho L, Hadden L, Connor P. Intake of Added Sugars and Selected Nutrients in the United States, National Health and Nutrition Examination Survey (NHANES) 2003—2006. Crit Rev Food Sci Nutr. 2010 Mar 8;50(3):228–58.

52. Centers for Disease Control and Prevention. Get the Facts: Sugar-Sweetened Beverages and Consumption [Internet]. 2018 [cited 2019 Oct 15]. Available from: https://www.cdc.gov/nutrition/data-statistics/sugar-sweetened-beverages-intake.html

53. Government of Canada SC. Change in total sugars consumption among Canadian children and adults [Internet]. 2019 [cited 2019 May 10]. Available from: https://www150.statcan.gc.ca/n1/pub/82-003-x/2019001/article/00002/tbl/tbl04-eng.htm

54. Singh GM, Micha R, Khatibzadeh S, Lim S, Ezzati M, Mozaffarian D. Estimated global, regional, and national disease burdens related to sugar-sweetened beverage consumption in 2010. Circulation. 2015;132(8):639–66.

55. Lythgoe A, Roberts C, Madden AM, Rennie KL. Marketing foods to children: a comparison of nutrient content between children's and non-children's products. Public Health Nutr. 2013 Dec;16(12):2221–30.

56. Fisberg M, Machado R. History of yogurt and current patterns of consumption. Nutr Rev. 2015 Aug 1;73(suppl_1):4–7.

57. Fernando J. Yogurt market: current status and consumption trends. Economics and Competitiveness Division, Alberta Agriculture and Rural Development. 2013 Oct;(23):5.

58. Brown-Riggs C. Nutrition and Health Disparities: The Role of Dairy in Improving Minority Health Outcomes. Int J Environ Res Public Health. 2016 Jan;13(1):1–9.

59. Statista. Topic: Yogurt market in Canada [Internet]. www.statista.com. 2019 [cited 2019 May 9]. Available from: https://www.statista.com/topics/4502/yogurt-market-in-canada/

60. Mozaffarian Dariush. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity. Circulation. 2016 Jan 12;133(2):187–225.

61. Marette A, Picard-Deland E. Yogurt consumption and impact on health: focus on children and cardiometabolic risk. Am J Clin Nutr. 2014 May 1;99(5):1243S-1247S.

62. Chollet M, Gille D, Schmid A, Walther B, Piccinali P. Acceptance of sugar reduction in flavored yogurt. J Dairy Sci. 2013 Sep 1;96(9):5501–11.

63. Moore JB, Horti A, Fielding BA. Evaluation of the nutrient content of yogurts: a comprehensive survey of yogurt products in the major UK supermarkets. BMJ Open. 2018 Aug;8:e021387.

64. Imamura F, Micha R, Khatibzadeh S, Fahimi S, Shi P, Powles J, et al. Dietary quality among men and women in 187 countries in 1990 and 2010: a systematic assessment. Lancet Glob Health. 2015 Mar 1;3(3):e132–42.

65. World Health Organization. Double burden of malnutrition [Internet]. [cited 2019 Sep 18]. Available from: http://www.who.int/nutrition/double-burden-malnutrition/en/

66. Nshimyumukiza L, Lieffers JRL, Ekwaru JP, Ohinmaa A, Veugelers PJ. Temporal changes in diet quality and the associated economic burden in Canada. PLOS ONE. 2018 Nov 8;13(11):e0206877.

67. Tugault-Lafleur CN, Black JL. Differences in the Quantity and Types of Foods and Beverages Consumed by Canadians between 2004 and 2015. Nutrients. 2019 Mar;11(3):526.

68. Brug J. Determinants of healthy eating: motivation, abilities and environmental opportunities. Fam Pract. 2008 Dec 1;25(suppl 1):i50–5.

69. Shavers VL. Measurement of socioeconomic status in health disparities research. J Natl Med Assoc. 2007 Sep;99(9):1013–23.

70. Kirkpatrick S, Tarasuk V. The relationship between low income and household food expenditure patterns in Canada. Public Health Nutr. 2003 Sep;6(6):589–97.

71. Ricciuto L, Tarasuk V, Yatchew A. Socio-demographic influences on food purchasing among Canadian households. Eur J Clin Nutr. 2006 Jun;60(6):778–90.

72. Dumitrescu C, Hughner RS, II CJS. Why Do Some Consumers Make Unhealthy Food Choices? J Food Prod Mark. 2016 May 4;0(0):1–22.

73. Popkin BM, Zizza C, Siega-Riz AM. Who is leading the change?: U.S. dietary quality comparison between 1965 and 1996. Am J Prev Med. 2003 Jul;25(1):1–8.

74. Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. Am J Public Health. 1992 Jun;82(6):816–20.

75. Azagba S, Sharaf MF. Disparities in the frequency of fruit and vegetable consumption by socio-demographic and lifestyle characteristics in Canada. Nutr J. 2011;10:118.

76. Sanou D, O'Reilly E, Ngnie-Teta I, Batal M, Mondain N, Andrew C, et al. Acculturation and Nutritional Health of Immigrants in Canada: A Scoping Review. J Immigr Minor Health. 2014;16(1):24–34.

77. Turrell G, Vandevijvere S. Socio-economic inequalities in diet and body weight: evidence, causes and intervention options. Public Health Nutr. 2015 Apr;18(05):759–763.

78. 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 Mar 1;16(2):175–90.

79. Oakes JM, Andrade KE, Biyoow IM, Cowan LT. Twenty Years of Neighborhood Effect Research: An Assessment. Curr Epidemiol Rep. 2015 Mar;2(1):80–7.

80. Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. Obes Rev Off J Int Assoc Study Obes. 2011 Mar;12(3):217–30.

81. Papas MA, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC. The Built Environment and Obesity. Epidemiol Rev. 2007 Jan 1;29(1):129–43.

82. Toms R, Bonney A, Mayne DJ, Feng X, Walsan R. Geographic and area-level socioeconomic variation in cardiometabolic risk factor distribution: a systematic review of the literature. Int J Health Geogr. 2019 Dec;18(1):149–60.

83. Ford PB, Dzewaltowski DA. Disparities in obesity prevalence due to variation in the retail food environment: three testable hypotheses. Nutr Rev. 2008 Apr 1;66(4):216–28.

84. Bhakat RS, Muruganantham G. A Review of Impulse Buying Behavior. Int J Mark Stud [Internet]. 2013 Apr 22 [cited 2015 Jul 7];5(3). Available from: http://www.ccsenet.org/journal/index.php/ijms/article/view/24570

85. Cohen DA, Collins R, Hunter G, Ghosh-Dastidar B, Dubowitz T. Store Impulse Marketing Strategies and Body Mass Index. Am J Public Health. 2014 Dec 18;105(7):1446–52.

86. Cairns G, Angus K, Hastings G, Caraher M. Systematic reviews of the evidence on the nature, extent and effects of food marketing to children. A retrospective summary. Appetite. 2013 Mar 1;62:209–15.

87. Cohen DA, Lesser LI. Obesity prevention at the point of purchase. Obes Rev. 2016 Jan 1;17(5):389–396.

88. 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 Promot Chronic Dis Prev Can Res Policy Pract. 2017 Oct;37(10):357–62.

89. Centers for Disease Control and Prevention. Census Tract Level State Maps of the Modified Retail Food Environment Index (mRFEI) [Internet]. 2011 [cited 2018 Mar 4]. Available from: https://www.cdc.gov/obesity/downloads/census-tract-level-state-maps-mrfei_TAG508.pdf

90. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: A systematic review. Health Place. 2012 Sep;18(5):1172–87.

91. Heart and Stroke Foundation Canada. Access to Affordable, healthy and nutritions food, Position statement [Internet]. 2017. Available from: https://www.heartandstroke.ca/-/media/pdf-files/canada/2017-position-statements/accessto-affordablehealthy-nutritiousfoods-ps-eng.ashx?la=en

92. Beatrice AB, Elizabeth M, Meaghan RB, Lynn R, Rebecca T. The Ontario Food and Nutrition Strategy: identifying indicators of food access and food literacy for early monitoring of the food environment. Health Promot Chronic Dis Prev Can Res Policy Pract. 2017 Sep;37(9):313–9.

93. Government of Canada, Health Canada. Measuring the Food Environment in Canada – Health Canada. 2013 Oct 9 [cited 2014 Dec 12]; Available from: http://www.hc-sc.gc.ca/fn-an/nutrition/pol/som-ex-sum-environ-eng.php

94. Stevenson AC, Brazeau A-S, Dasgupta K, Ross NA. Neighbourhood retail food outlet access, diet and body mass index in Canada: a systematic review. Health Promot Chronic Dis Prev Can Res Policy Pract. 2019 Oct;39(10):261–80.

95. Beaulac J, Kristjansson E, Cummins S. A Systematic Review of Food Deserts, 1966-2007. Prev Chronic Dis. 2009 Jun 15;6(3):A 105.

96. Cummins S, Macintyre S. “Food deserts”—evidence and assumption in health policy making. BMJ. 2002 Aug 24;325(7361):436–8.

97. Vaughan CA, Cohen DA, Ghosh-Dastidar M, Hunter GP, Dubowitz T. Where do food desert residents buy most of their junk food? Supermarkets. Public Health Nutr. 2016 Oct 5;20(14):2608–16.

98. Lamb KE, Thornton LE, Cerin E, Ball K. Statistical approaches used to assess the equity of access to food outlets: A systematic review. Aimsph 2015 Vol 2 Pages 358-401. 2015 Jul 19;2(3):358–401.

99. Hawkes C. Sales promotions and food consumption. Nutr Rev. 2009 Jun 1;67(6):333–42.

100. Ailawadi KL, Neslin SA. The Effect of Promotion on Consumption: Buying More and Consuming It Faster. J Mark Res. 1998;35(3):390–8.

101. Black C, Moon G, Baird J. Dietary inequalities: What is the evidence for the effect of the neighbourhood food environment? Health Place. 2014 May 1;27:229–42.

102. Francesca Colantuoni. Empirical Studies On The Effectiveness Of Soda Taxes To Curb Obesity - viewcontent.cgi [Internet]. [MASSACHUSETTS]: University of Massachusetts Amherst; 2014 [cited 2016 Jan 15]. Available from: http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1118&context=dissertations_2

103. Choi K, Hennrikus D, Forster J, St. Claire AW. Use of Price-Minimizing Strategies by Smokers and Their Effects on Subsequent Smoking Behaviors. Nicotine Tob Res. 2012 Jul 1;14(7):864–70.

104. Zorbas C, Eyles H, Orellana L, Peeters A, Mhurchu CN, Riesenberg D, et al. Do purchases of price promoted and generic branded foods and beverages vary according to food category and income level? Evidence from a consumer research panel. Appetite. 2019 Oct 4;104481.

105. Marmot M, Friel S, Bell R, Houweling TA, Taylor S. Closing the gap in a generation: health equity through action on the social determinants of health. The Lancet. 2008 Nov 8;372(9650):1661–9.

106. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet. 2014 Jul 5;384(9937):45–52.

107. Drewnowski A. Obesity, diets, and social inequalities. Nutr Rev. 2009 May 1;67(suppl_1):S36–9.

108. Singh GK, Kogan MD, Van Dyck PC, Siahpush M. Racial/Ethnic, Socioeconomic, and Behavioral Determinants of Childhood and Adolescent Obesity in the United States: Analyzing Independent and Joint Associations. Ann Epidemiol. 2008 Sep 1;18(9):682–95.

109. Vogel C, Ntani G, Inskip H, Barker M, Cummins S, Cooper C, et al. Education and the Relationship Between Supermarket Environment and Diet. Am J Prev Med. 2016 Aug;51(2):e27-e34.

110. Dendup T, Feng X, Clingan S, Astell-Burt T. Environmental Risk Factors for Developing Type 2 Diabetes Mellitus: A Systematic Review. Int J Environ Res Public Health. 2018 Jan 5;15(1).

111. Acevedo-Garcia D, Osypuk TL, McArdle N, Williams DR. Toward A Policy-Relevant Analysis Of Geographic And Racial/Ethnic Disparities In Child Health. Health Aff (Millwood). 2008 Mar 1;27(2):321–33.

112. Alageel S, Gulliford MC, McDermott L, Wright AJ. Multiple health behaviour change interventions for primary prevention of cardiovascular disease in primary care: systematic review and meta-analysis. BMJ Open. 2017 Jun 1;7(6):e015375.

113. Moore L, de Silva-Sanigorski A, Moore SN. A socio-ecological perspective on behavioural interventions to influence food choice in schools: alternative, complementary or synergistic? Public Health Nutr. 2013 Jun;16(6):1000–5.

114. Frieden TR. A Framework for Public Health Action: The Health Impact Pyramid. Am J Public Health. 2010 Apr 1;100(4):590–5.

115. McGill R, Anwar E, Orton L, Bromley H, Lloyd-Williams F, O'Flaherty M, et al. Are interventions to promote healthy eating equally effective for all? Systematic review of socioeconomic inequalities in impact. BMC Public Health [Internet]. 2015 May 2 [cited 2015 Sep 20];15(457). Available from: http://www.biomedcentral.com/1471-2458/15/457/abstract

116. Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. Prev Med. 1999 Dec;29(6 Pt 1):563–70.

117. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. Am J Public Health. 1998 Feb 1;88(2):216–22.

118. Engler-Stringer R, Le H, Gerrard A, Muhajarine N. The community and consumer food environment and children's diet: a systematic review. BMC Public Health. 2014 May 29;14(522):1–15.

119. Bes-Rastrollo M, Sayon-Orea C, Ruiz-Canela M, Martinez-Gonzalez MA. Impact of sugars and sugar taxation on body weight control: A comprehensive literature review. Obesity. 2016 Jul 1;24(7):1410–26.

120. Anne Marie Thow, Downs S, Jan S. A systematic review of the effectiveness of food taxes and subsidies to improve diets: Understanding the recent evidence. Nutr Rev. 2014 Sep;72(9):551–65.

121. Waterlander WE, Jiang Y, Nghiem N, Eyles H, Wilson N, Cleghorn C, et al. The effect of food price changes on consumer purchases: a randomised experiment. Lancet Public Health. 2019 Aug 1;4(8):e394–405.

122. Vanderlee L, Goorang S, Karbasy K, Schermel A, L'Abbe M. Creating healthier food environments in Canada: Current policies and priority actions - Summary report. [Internet]. Toronto: University of Toronto; 2017. Available from: http://labbelab.utoronto.ca/Food-EPI-Canada-2017/

123. Christoforou A, Trieu K, Land M-A, Bolam B, Webster J. State-level and community-level salt reduction initiatives: a systematic review of global programmes and their impact. J Epidemiol Community Health. 2016 Nov 1;70(11):1140–50.

124. Pratt CA, Loria CM, Arteaga SS, Nicastro HL, Lopez-Class M, de Jesus JM, et al. A Systematic Review of Obesity Disparities Research. Am J Prev Med. 2017 Jul;53(1):113–22.

125. Stokols D. Establishing and maintaining healthy environments. Toward a social ecology of health promotion. Am Psychol. 1992 Jan;47(1):6–22.

126. Sallis JF, Owen N, Fisher EB. Ecological models of health behavior. In: Glanz K, Rimer BK, Viswanath K, editors. Health behavior and health education: theory, research, and practice. 4th ed. San Francisco: Jossey-Bass; 2008. p. 465–552.

127. Schwartz MB, Schneider GE, Choi Y-Y, Li X, Harris J, Andreyeva T, et al. Association of a Community Campaign for Better Beverage Choices With Beverage Purchases From Supermarkets. JAMA Intern Med. 2017 May 1;177(5):666–74.

128. The CDC guide to strategies for reducing the consumption of sugar-sweetened beverages [Internet]. 2010 Mar. Available from: https://stacks.cdc.gov/view/cdc/51532

129. Jennings A, Cassidy A, Winters T, Barnes S, Lipp A, Holland R, et al. Positive effect of a targeted intervention to improve access and availability of fruit and vegetables in an area of deprivation. Health Place. 2012 Sep 1;18(5):1074–8.

130. Barnhill A, Palmer A, Weston CM, Brownell KD, Clancy K, Economos CD, et al. Grappling With Complex Food Systems to Reduce Obesity: A US Public Health Challenge. Public Health Rep. 2018 Nov 1;133(1_suppl):44S-53S.

131. Thacker SB, Berkelman RL. Public Health Surveillance in the United States. Epidemiol Rev. 1988 Jan 1;10(1):164–90.

132. Seliske L, Norwood TA, McLaughlin JR, Wang S, Palleschi C, Holowaty E. Estimating micro area behavioural risk factor prevalence from large population-based surveys: a full Bayesian approach. BMC Public Health. 2016 Jun 7;16(1):478.

133. Cheadle A, Psaty BM, Diehr P, Koepsell T, Wagner E, Curry S, et al. Evaluating Community-Based Nutrition-Programs: Comparing Grocery Store and Individual-Level Survey Measures of Program Impact. Prev Med. 1995 Jan 1;24(1):71–9.

134. Cheadle A, Sterling TD, Schmid TL, Fawcett SB. Promising community-level indicators for evaluating cardiovascular health-promotion programs. Health Educ Res. 2000 Feb;15(1):109–16.

135. Abarca J, Ramachandran S. Using community indicators to assess nutrition in Arizona-Mexico border communities. Prev Chronic Dis. 2005 Jan;2(1):A06.

136. Brown AF, Ma GX, Miranda J, Eng E, Castille D, Brockie T, et al. Structural Interventions to Reduce and Eliminate Health Disparities. Am J Public Health. 2019 Jan 1;109(S1):S72–8.

137. Leeflang PSH, Bijmolt THA, van Doorn J, Hanssens DM, van Heerde HJ, Verhoef PC, et al. Creating lift versus building the base: Current trends in marketing dynamics. Int J Res Mark. 2009 Mar 1;26(1):13–20.

138. Leeflang P, Wieringa JE, Bijmolt THA, Pauwels KH. Modeling Markets: Analyzing Marketing Phenomena and Improving Marketing Decision Making. New York: Springer-Verlag; 2015. (International Series in Quantitative Marketing).

139. Government of Canada, Statistics Canada. 2011 National Household Survey Profile - Census metropolitan area/Census agglomeration [Internet]. 2013 [cited 2016 May 16]. Available from: https://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMA&Code1=462&Data=Count&SearchText=462&SearchType=Begins&SearchPR=01&A1=All&B1=All&Custom=&TABID=3

140. Statistics Canada. Census metropolitan area (CMA) and census agglomeration (CA) - Census Dictionary [Internet]. [cited 2019 Aug 4]. Available from: https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo009-eng.cfm

141. Hollands S, Campbell MK, Gilliland J, Sarma S. A spatial analysis of the association between restaurant density and body mass index in Canadian adults. Prev Med. 2013 Oct;57(4):258–64.

142. Minaker LM, McCargar L, Lambraki I, Jessup L, Driezen P, Calengor K, et al. School Region Socio-economic Status and Geographic Locale is Associated with Food Behaviour of Ontario and Alberta Adolescents. Can J Public Health Rev Can Santee Publique. 2006;97(5):357–61.

143. Statistics Canada. Postal Code Conversion File, Reference Guide [Internet]. 2011 [cited 2016 Dec 8]. Available from: http://www.statcan.gc.ca/pub/92-153-g/92-153-g2011002-eng.htm

144. Ross NA, Tremblay S, Graham K. Neighbourhood influences on health in Montréal, Canada. Soc Sci Med. 2004 Oct;59(7):1485–94.

145. Statistics Canada. Dissemination area (DA) - census dictionary [Internet]. [cited 2019 May 23]. Available from: https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo021-eng.cfm

146. Public Health Agency of Canada. Diabetes in Canada [Internet]. aem. 2017 [cited 2019 Oct 22]. Available from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/diabetes-canada-highlights-chronic-disease-surveillance-system.html

147. Pitney Bowes Canada. Product documentation: Canada business data. 2012.

148. Daepp MI, Black J. Assessing the validity of commercial and municipal food environment data sets in Vancouver, Canada. Public Health Nutr. 2017 Oct;20(15):2649–59.

149. Clayton D, Bernardinelli L. Bayesian methods for mapping disease risk. In: Elliot P, Cuzik D, English D, Stern R, editors. Geographical and Environmental Epidemiology: Methods for Small Area Studies. Oxford University Press; 1996.

150. Elliott P, Wakefield JC, Best NG, Briggs DJ. Spatial epidemiology: methods and applications. Oxford University Press; 2001.

151. Besag J. Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Ser B Methodol. 1974;36(2):192–236.

152. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math. 1991 Mar 1;43(1):1–20.

153. Clayton DG, Bernardinelli L, Montomoli C. Spatial Correlation in Ecological Analysis. Int J Epidemiol. 1993 Dec 1;22(6):1193–202.

154. Hodges JS, Reich BJ. Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love. Am Stat. 2010 Nov 1;64(4):325–34.

155. Reich BJ, Hodges JS, Zadnik V. Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models. Biometrics. 2006;62(4):1197–206.

156. Hughes J, Haran M. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. J R Stat Soc Ser B Stat Methodol. 2013 Jan 1;75(1):139–59.

157. Banerjee S, Carlin B, Alan G. Hierarchical Modeling and Analysis for Spatial Data, Second Edition. 2nd ed. Boca Raton, FL: CRC Press; 2014.

158. Bernardinelli L, Clayton D, Montomoli C. Bayesian estimates of disease maps: how important are priors? Stat Med. 1995 Nov 15;14(21–22):2411–31.

159. Leroux BG, Lei X, Breslow N. Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence. In: Halloran ME, Berry D, editors. Statistical Models in Epidemiology, the Environment, and Clinical Trials. New York, NY: Springer; 2000. p. 179–91. (The IMA Volumes in Mathematics and its Applications).

160. Riebler A, Sørbye SH, Simpson D, Rue H. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. ArXiv160101180 Stat [Internet]. 2016 Jan 6 [cited 2019 May 6]; Available from: http://arxiv.org/abs/1601.01180

161. Friske W, Choi S. Another Look at Retail Gravitation Theory: History, Analysis, and Future Considerations. ABD J [Internet]. 2013;5(1). Available from:

https://www.ship.edu/contentassets/569211b0c6f243808c3c64f54e816cd2/another-look-at-retail-gravitation-theory.pdf

162. Anderson SJ, Volker JX, Phillips MD. Converse's Breaking-Point Model Revised. J Manag Mark Res. 2010 Jan;3:1.

163. Reilly WJ. The law of retail gravitation. New York: Knickerbocker Press; 1931.

164. Drezner T. A review of competitive facility location in the plane. Logist Res. 2014 Jun 4;7(1):114.

165. Kant AK, Graubard BI. Secular trends in the association of socio-economic position with self-reported dietary attributes and biomarkers in the US population: National Health and Nutrition Examination Survey (NHANES) 1971–1975 to NHANES 1999–2002. Public Health Nutr. 2007 Feb;10(2):158–67.

166. Institute for Health Metrics and Evaluation. Global Burden of Disease (GBD) - United States [Internet]. Global Burden of Disease (GBD) Country Profile. [cited 2016 Sep 9]. Available from: http://www.healthdata.org/united-states

167. World Health Organization. Obesity and overweight [Internet]. WHO Fact Sheets. 2018 [cited 2015 Dec 23]. Available from: http://www.who.int/mediacentre/factsheets/fs311/en/

168. Federal Trade Commission. Review of Food Marketing to Children and Adolescents -- Follow-Up Report [Internet]. 2012 [cited 2017 Jan 3]. Available from: https://www.ftc.gov/reports/review-food-marketing-children-adolescents-follow-report

169. Welsh JA, Sharma AJ, Grellinger L, Vos MB. Consumption of added sugars is decreasing in the United States. Am J Clin Nutr. 2011 Sep 1;94(3):726–34.

170. 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. In: Preedy VR, Hunter L-A, Patel VB, editors. Diet Quality. Springer New York; 2013. p. 383–96. (Nutrition and Health).

171. Mela CF, Gupta S, Lehmann DR. The Long-Term Impact of Promotion and Advertising on Consumer Brand Choice. J Mark Res. 1997;34(2):248–61.

172. Neslin SA, Henderson C, Quelch J. Consumer Promotions and the Acceleration of Product Purchases. Mark Sci. 1985;4(2):147–65.

173. Bell DR, Chiang J, Padmanabhan V. The Decomposition of Promotional Response: An Empirical Generalization. Mark Sci. 1999 Nov;18(4):504–26.

174. James WP, Nelson M, Ralph A, Leather S. Socioeconomic determinants of health. The contribution of nutrition to inequalities in health. BMJ. 1997 May 24;314(7093):1545–9.

175. Centers for Disease Control and Prevention. CDC Health Disparities and Inequalities Report — United States, 2013. 2013 Nov 22;62(3):1–187.

176. Nielsen. Retail Food and Beverage. Nielsen Market Track, Quebec Grocery Banners, 52 Weeks, Jan - Dec 2012. 2013.

177. Ricciuto L, Tarasuk V, Yatchew A. Socio-demographic influences on food purchasing among Canadian households. Eur J Clin Nutr. 2006 Jun;60(6):778–90.

178. Godley J, Mclaren L. Socioeconomic Status and Body Mass Index in Canada: Exploring Measures and Mechanisms. Can Rev Sociol Can Sociol. 2010 Nov 1;47(4):381–403.

179. Ross NA, Tremblay S, Khan S, Crouse D, Tremblay M, Berthelot J-M. Body Mass Index in Urban Canada: Neighborhood and Metropolitan Area Effects. Am J Public Health. 2007 Mar;97(3):500–8.

180. Daniel M, Kestens Y, Paquet C. Demographic and urban form correlates of healthful and unhealthful food availability in Montréal, Canada. Can J Public Health. 2009;100(3):189–93.

181. Hu FB. Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. Obes Rev Off J Int Assoc Study Obes. 2013 Aug;14(8):606–19.

182. Government of Canada, Statistics Canada. Statistics Canada: 2011 Census Profile [Internet]. 2012 [cited 2015 Dec 29]. Available from: https://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMA&Code1=462&Geo2=PR&Code2=01&Data=Count&SearchText=montreal&SearchType=Begins&SearchPR=24&B1=All&Custom=&TABID=1

183. Government of Canada, Statistics Canada. Industry classifications [Internet]. 2011 [cited 2016 Feb 17]. Available from: http://www.statcan.gc.ca/eng/concepts/industry

184. Trivedi M, Gauri DK, Ma Y. Measuring the Efficiency of Category-Level Sales Response to Promotions. Manag Sci. 2016 Sep 1;63(10):3473–88.

185. Ma S, Fildes R. A retail store SKU promotions optimization model for category multi-period profit maximization. Eur J Oper Res. 2017 Jul 16;260(2):680–92.

186. Raju JS. The Effect of Price Promotions on Variability in Product Category Sales. Mark Sci. 1992;11(3):207–20.

187. Nijs VR, Dekimpe MG, Steenkamps J-BEM, Hanssens DM. The Category-Demand Effects of Price Promotions. Mark Sci. 2000 Dec;20(1):1–22.

188. Fox EJ, Montgomery AL, Lodish LM. Consumer Shopping and Spending across Retail Formats. J Bus. 2004;77(S2):S25–60.

189. Ng SW, Popkin BM. Monitoring foods and nutrients sold and consumed in the United States: Dynamics and Challenges. J Acad Nutr Diet. 2012 Jan;112(1):41-45.ed.

190. Vaughan CA, Collins R, Ghosh-Dastidar M, Beckman R, Dubowitz T. Does where you shop or who you are predict what you eat?: The role of stores and individual characteristics in dietary intake. Prev Med. 2017 Jul;100(12):10–6.

191. Gauri DK, Sudhir K, Talukdar D. The Temporal and Spatial Dimensions of Price Search: Insights from Matching Household Survey and Purchase Data. J Mark Res. 2008;45(2):226–40.

192. Urbany JE, Dickson PR, Sawyer AG. Insights into cross- and within-store price search: retailer estimates vs. consumer self-reports. J Retail. 2000 Jun 1;76(2):243–58.

193. Finkelstein EA, Strombotne KL, Zhen C, Epstein LH. Food Prices and Obesity: A Review. Adv Nutr Int Rev J. 2014 Nov 1;5(6):818–21.

194. Eisenberg MJ, Atallah R, Grandi SM, Windle SB, Berry EM. Legislative approaches to tackling the obesity epidemic. Can Med Assoc J. 2011 Sep 20;183(13):1496–500.

195. Stern D, Robinson WR, Ng SW, Gordon-Larsen P, Popkin BM. US Household Food Shopping Patterns: Dynamic Shifts Since 2000 And Socioeconomic Predictors. Health Aff (Millwood). 2015 Nov 1;34(11):1840–8.

196. Houle SKD, Rosenthal MM, Campbell NRC, Duhaney T, Tsuyuki RT. Why pharmacists should care about the marketing of unhealthy foods: Increasing our role in public health policy / Pourquoi les pharmaciens devraient-ils s'intéresser à la promotion des aliments néfastes pour la santé? Contribuons davantage aux politiques en matière de santé publique. Can Pharm J CPJ. 2014 Jan;147(1):4–7.

197. Li C, Balluz LS, Okoro CA, Strine TW, Lin J-MS, Town M, et al. Surveillance of Certain Health Behaviors and Conditions Among States and Selected Local Areas --- Behavioral Risk Factor Surveillance System, United States, 2009. Morb Mortal Wkly Rep. 2011 Aug 19;60(SS09):1–250.

198. Khan LK, Sobush K, Keener D, Goodman K, Lowry A, Kakietek J, et al. Recommended community strategies and measurements to prevent obesity in the United States. MMWR Recomm Rep Morb Mortal Wkly Rep Recomm Rep Cent Dis Control. 2009 Jul 24;58(RR-7):1–26.

199. Myers CA, Slack T, Martin CK, Broyles ST, Heymsfield SB. Regional Disparities in Obesity Prevalence in the United States: A Spatial Regime Analysis. Obes Silver Spring Md. 2015 Feb;23(2):481–7.

200. Hipp JA. Spatial Analysis and Correlates of County-Level Diabetes Prevalence, 2009 - 2010. Prev Chronic Dis. 2015 Jan 22;12.

201. University of Wisconsin Population Health Institute, Robert Wood Johnson Foundation. How Healthy is your County? | County Health Rankings [Internet]. County Health Rankings & Roadmaps. 2016 [cited 2018 Dec 17]. Available from: http://www.countyhealthrankings.org/explore-health-rankings/what-and-why-we-rank/health-factors/health-behaviors/diet-and-exercise

202. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CAM. The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results. Obes Silver Spring Md. 2015 Jul;23(7):1331–44.

203. Duffey KJ, Popkin BM. Adults with healthier dietary patterns have healthier beverage patterns. J Nutr. 2006;136(11):2901–7.

204. Frary CD, Johnson RK, Wang MQ. Children and adolescents' choices of foods and beverages high in added sugars are associated with intakes of key nutrients and food groups. J Adolesc Health. 2004 Jan;34(1):56–63.

205. National Research Council. Proprietary Data Sources. In: Improving Data to Analyze Food and Nutrition Policies. Washington, DC: National Academic Press; 2005. p. 53–62.

206. Statistics Canada, Government of Canada. North American Industry Classification System (NAICS) Canada 2012 [Internet]. 2011 [cited 2015 Dec 30]. Available from: http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=118464

207. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique and future directions. Stat Med. 2009;28(25):3049–67.

208. Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. PLOS ONE. 2013 Oct 9;8(10):e75256.

209. Wakefield J. Disease mapping and spatial regression with count data. Biostatistics. 2007 Apr 1;8(2):158–83.

210. Hanks EM, Schliep EM, Hooten MB, Hoeting JA. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. Environmetrics. 2015 Jun 1;26(4):243–54.

211. John Hughes. ngspatial: A Package for Fitting the Centered Autologistic and Sparse Spatial Generalized Linear Mixed Models for Areal Data. R J. 2015 Jan 4;6(2):81–95.

212. Spiegelhalter DJ, Best NG, Carlin BP. Bayesian measures of model complexity and fit. J R Stat Soc Ser B. 2002;64(4):583–639.

213. Stern D, Poti JM, Ng SW, Robinson WR, Gordon-Larsen P, Popkin BM. Where people shop is not associated with the nutrient quality of packaged foods for any racial-ethnic group in the United States. Am J Clin Nutr. 2016 Apr 1;103(4):1125–34.

214. Direction régionale de santé publique de Montréal. Atlas Santé Montréal - Maladies chroniques [Internet]. 2011 [cited 2018 May 8]. Available from: http://www.cmisatlas.mtl.rtss.qc.ca/geoclip_dyn/mal_chron/carto.php?lang=fr&nivgeos=clsc&curCodeDomCH=Mpoc&curCodeThemeCH=MpocPrev&typindCH=C&curCodeIndCH=t_35_plus&curserieCH=2009-10

215. Drewnowski A, Rehm CD, Moudon AV, Arterburn D. The Geography of Diabetes by Census Tract in a Large Sample of Insured Adults in King County, Washington, 2005–2006. Prev Chronic Dis. 2014 Jul 24;11:E125.

216. Congdon P. Variations in Obesity Rates between US Counties: Impacts of Activity Access, Food Environments, and Settlement Patterns. Int J Environ Res Public Health. 2017 07;14(9):E1023.

217. Timmins KA, Green MA, Radley D, Morris MA, Pearce J. How has big data contributed to obesity research? A review of the literature. Int J Obes. 2018 Jul 18;(42):1952–62.

218. Hruby A, Hu FB. The Epidemiology of Obesity: A Big Picture. PharmacoEconomics. 2015 Jul;33(7):673–89.

219. Richards CL, Iademarco MF, Atkinson D, Pinner RW, Yoon P, Mac Kenzie WR, et al. Advances in Public Health Surveillance and Information Dissemination at the Centers for Disease Control and Prevention. Public Health Rep. 2017 Jul 1;132(4):403–10.

220. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2019 May 11;393(10184):1958–72.

221. Mamiya H, Schmidt AM, Moodie EEM, Ma Y, Buckeridge DL. An Area-Level Indicator of Latent Soda Demand: Spatial Statistical Modeling of Grocery Store Transaction Data to Characterize the Nutritional Landscape in Montreal, Canada. Am J Epidemiol. 2019 Sep;188(9):1713–1722.

222. Huff DL. Defining and estimating a trading area. J Mark. 1964;28(3):34–38.

223. Narain A, Kwok CS, Mamas MA. Soft drinks and sweetened beverages and the risk of cardiovascular disease and mortality: a systematic review and meta-analysis. Int J Clin Pract. 2016;70(10):791–805.

224. Toews I, Lohner S, Gaudry DK de, Sommer H, Meerpohl JJ. Association between intake of non-sugar sweeteners and health outcomes: systematic review and meta-analyses of randomised and non-randomised controlled trials and observational studies. BMJ. 2019 Jan 2;364:k4718.

225. Sylvetsky AC, Rother KI. Trends in the Consumption of Low-Calorie Sweeteners. Physiol Behav. 2016 Oct 1;164(Pt B):446–50.

226. Evaluation of the nutrient content of yogurts: a comprehensive survey of yogurt products in the major UK supermarkets | BMJ Open [Internet]. [cited 2019 May 14]. Available from: https://bmjopen.bmj.com/content/8/8/e021387

227. Hillier A, Smith TE, Whiteman ED, Chrisinger BW. Discrete choice model of food store trips using national household food acquisition and purchase survey (FoodAPS). Int J Environ Res Public Health. 2017 Sep 27;14(10):1133.

228. Government of Canada, Statistics Canada. Road network file, 2013 [Internet]. 2013 [cited 2019 May 21]. Available from: https://www150.statcan.gc.ca/n1/en/catalogue/92-500-X2013001

229. Nakanishi M, Cooper LG. Parameter estimation for a multiplicative competitive interaction model: least squares approach. J Mark Res. 1974;11(3):303–11.

230. Huff D, McCallum BM. Calibrating the Huff Model Using ArcGIS Business Analyst [Internet]. ESRI; 2008 [cited 2019 Sep 9]. Available from: https://www.esri.com/library/whitepapers/pdfs/calibrating-huff-model.pdf

231. Pohjanheimo T, Sandell M. Explaining the liking for drinking yoghurt: The role of sensory quality, food choice motives, health concern and product information. Int Dairy J. 2009 Aug 1;19(8):459–66.

232. Howard Wilsher S, Harrison F, Yamoah F, Fearne A, Jones A. The relationship between unhealthy food sales, socio-economic deprivation and childhood weight status: results of a cross-sectional study in England. Int J Behav Nutr Phys Act. 2016 Feb 15;13(1):21.

233. Aiello LM, Schifanella R, Quercia D, Prete LD. Large-scale and high-resolution analysis of food purchases and health outcomes. EPJ Data Sci. 2019 Dec;8(1):14.

234. Black C, Ntani G, Inskip H, Cooper C, Cummins S, Moon G, et al. Measuring the healthfulness of food retail stores: variations by store type and neighbourhood deprivation. Int J Behav Nutr Phys Act. 2014 May 23;11(1):69.

235. Pechey R, Monsivais P. Supermarket Choice, Shopping Behavior, Socioeconomic Status, and Food Purchases. Am J Prev Med. 2015 Dec;49(6):868–77.

236. Chaix B, Bean K, Daniel M, Zenk SN, Kestens Y, Charreire H, et al. Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). PLoS ONE. 2012 Apr 4;7(4):e32908.

237. Krukowski RA, Sparks C, DiCarlo M, McSweeney J, West DS. There's more to food store choice than proximity: a questionnaire development study. BMC Public Health. 2013;13:586.

238. Nielsen. Understanding the top drivers behind shoppers' store choices [Internet]. 2016 [cited 2019 Jul 24]. Available from: https://www.nielsen.com/us/en/insights/article/2016/understanding-the-top-drivers-behind-shoppers-store-choices

239. Riesenberg D, Backholer K, Zorbas C, Sacks G, Paix A, Marshall J, et al. Price Promotions by Food Category and Product Healthiness in an Australian Supermarket Chain, 2017–2018. Am J Public Health. 2019 Aug 15;109(10):1434–9.

240. Bandy L, Adhikari V, Jebb S, Rayner M. The use of commercial food purchase data for public health nutrition research: A systematic review. PLOS ONE. 2019 Jan 7;14(1):e0210192.

241. Ataman MB, Van Heerde HJ, Mela CF. The Long-Term Effect of Marketing Strategy on Brand Sales. J Mark Res. 2010 Oct 1;47(5):866–82.

242. Anderson ET, Simester DI. Long-Run Effects of Promotion Depth on New Versus Established Customers: Three Field Studies. Mark Sci. 2004 Feb 1;23(1):4–20.

243. Kopalle PK, Mela CF, Marsh L. The Dynamic Effect of Discounting on Sales: Empirical Analysis and Normative Pricing Implications. Mark Sci. 1999 Aug 1;18(3):317–32.

244. Maher JH, Lowe J, Hughes R. Community pharmacy as a setting for public health nutrition action: Australian nutritionists' perspectives. Public Health Nutr. 2015 Jul;18(10):1864–72.

245. Centers for Disease Control and Prevention. Community Health Assessment [Internet]. 2019 [cited 2019 Aug 28]. Available from: https://www.cdc.gov/publichealthgateway/cha/data.html