

# Estimating spatial patterning of dietary behaviors using grocery transaction data

Hiroshi Mamiya\*, Erica Moodie and David Buckeridge

Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

## Objective

To demonstrate a method for estimating neighborhood food selection with secondary use of digital marketing data; grocery transaction records and retail business registry.

## Introduction

Unhealthy diet is becoming the most important preventable cause of chronic disease burden (1). Dietary patterns vary across neighborhoods as a function of policy, marketing, social support, economy, and the commercial food environment (2). Assessment of community-specific response to these socio-ecological factors is critical for the development and evaluation policy interventions and identification of nutrition inequality. Mass administration of dietary surveys is impractical and prohibitory expensive, and surveys typically fail to address variation of food selection at high geographic resolution. Marketing companies such as the Nielsen cooperation continuously collect and centralize scanned grocery transaction records from a geographically representative sample of retail food outlets to guide product promotions. These data can be harnessed to develop a model for the demand of specific foods using store and neighborhood attributes, providing a rich and detailed picture of the “foodscape” in an urban environment. In this study, we generated a spatial profile of food selection from estimated sales in food outlets in the Census Metropolitan Area (CMA) of Montreal, Canada, using regular carbonated soft drinks (i.e. non-diet soda) as an initial example.

## Methods

From the Nielsen cooperation, we obtained weekly grocery transaction data generated by a sample of 86 grocery stores and 42 pharmacies in the Montreal CMA in 2012. Extracted store-specific soda sales were standardized to a single serving size (240ml) and averaged across 52 weeks, resulting in 128 data points. Using linear regression, natural log-transformed soda sales were modelled as a function of store type (grocery vs. pharmacies), chain identification code and socio-demographic attributes of store neighborhood, which are median family income, proportion of individuals who received post-secondary diplomas, and population density as measured by the 2011 Canadian Household Survey. Selection of the predictors and first-order interaction terms was guided by the minimization of the mean squared error using 10-fold cross-validation. The final model was applied to all operating chain grocery stores and pharmacies in 2012 (n=980) recorded in a comprehensive and commonly available business establishment database. The resulting predicted store-specific weekly average soda sales was spatially interpolated to provide a graphical representation of the soda sales (representing an unhealthy foodscape) across the Montreal CMA.

## Results

Figure 2 demonstrates the spatial distribution of the predicted soda sales in the Montreal CMA.

## Conclusions

The current lack of neighborhood-level dietary surveillance impedes effective public health actions aimed at encouraging healthy

food selection and subsequent reduction of chronic illness. Our method leverages existing grocery transaction data and store location information to address the gap in population monitoring of nutrition status and urban foodscapes. Future applications of our methodology to other store types (e.g. convenience stores) and food products across multiple time points (e.g. months and years) will permit a comprehensive, timely and automated assessment of dietary trends, identification of neighborhoods in special dietary needs, development of tailored community health promotions, and the measurement of neighbourhood-specific response to nutrition policies and unhealthy food advertising.

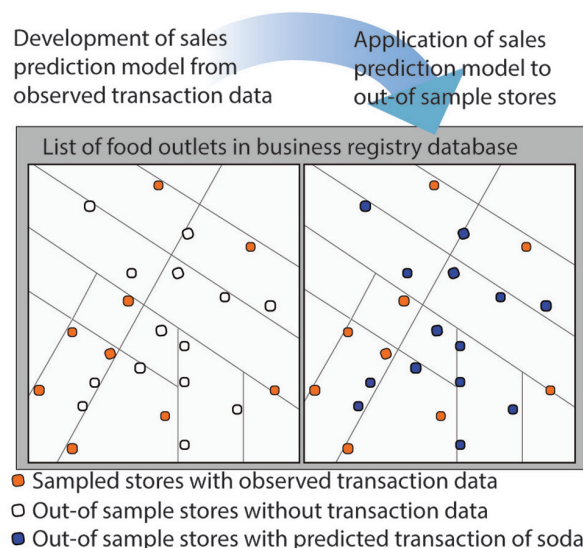


Figure 1: Schematic representation of the process generating spatial food selection measure using grocery transaction data and business establishment database

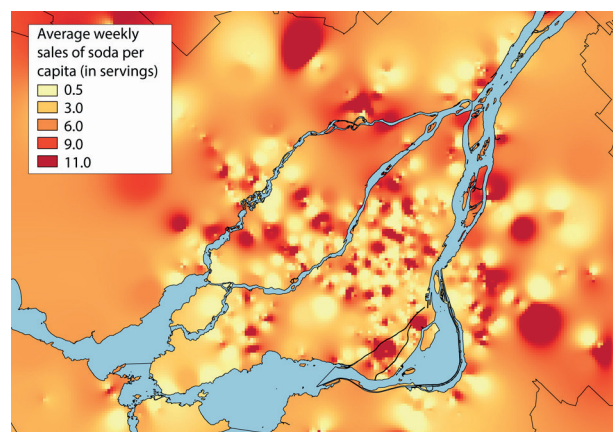


Figure 2: Predicted weekly sales of soda in the Montreal CMA in 2012. Spatial interpolation was performed on the point quantities of predicted sales at each store.



## ISDS 2016 Conference Abstracts

### Keywords

Chronic disease; Nutrition; Spatial analysis; Transaction data; Prediction

### Acknowledgments

We thank Dr. Luc De Montigny for providing location-based business registry.

### References

1. 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>
2. Richard L, Gauvin L, Raine K. Ecological Models Revisited: Their Uses and Evolution in Health Promotion Over Two Decades. *Annual Review of Public Health*. 2011;32(1):307–26.

### \*Hiroshi Mamiya

E-mail: [hiroshi.mamiya@mail.mcgill.ca](mailto:hiroshi.mamiya@mail.mcgill.ca)

