Public Health Nutrition: 24(17), 5616–5628 doi:10.1017/S1368980021003645

Generating community measures of food purchasing activities using store-level electronic grocery transaction records: an ecological study in Montreal, Canada

Hiroshi Mamiya[1,*], Alexandra M Schmidt[1], Erica EM Moodie[1], Yu Ma[2] and David L Buckeridge[1]

[1]School of Population and Global Health, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 772 Sherbrooke St West, Montreal, QC H3A 1G1, Canada: [2]Desautels Faculty of Management, McGill University, Montreal, QC, Canada

Submitted 1 March 2021: Final revision received 9 June 2021: Accepted 17 August 2021: First published online 23 August 2021

Abstract

Objective: Geographic measurement of diets is generally not available at areas smaller than a national or provincial (state) scale, as existing nutrition surveys cannot achieve sample sizes needed for an acceptable statistical precision for small geographic units such as city subdivisions.

Design: Using geocoded Nielsen grocery transaction data collected from supermarket, supercentre and pharmacy chains combined with a gravity model that transforms store-level sales into area-level purchasing, we developed small-area public health indicators of food purchasing for neighbourhood districts. We generated the area-level indicators measuring per-resident purchasing quantity for soda, diet soda, flavoured (sugar-added) yogurt and plain yogurt purchasing. We then provided an illustrative public health application of these indicators as covariates for an ecological spatial regression model to estimate spatially correlated small-area risk of type 2 diabetes mellitus (T2D) obtained from the public health administrative data.

Setting: Greater Montreal, Canada in 2012.

Participants: Neighbourhood districts (*n* 193).

Results: The indicator of flavoured yogurt had a positive association with neighbourhood-level risk of T2D (1·08, 95 % credible interval (CI) 1·02, 1·14), while that of plain yogurt had a negative association (0·93, 95 % CI 0·89, 0·96). The indicator of soda had an inconclusive association, and that of diet soda was excluded due to collinearity with soda. The addition of the indicators also improved model fit of the T2D spatial regression (Watanabe–Akaike information criterion = 1765 with the indicators, 1772 without).

Conclusion: Store-level grocery sales data can be used to reveal micro-scale geographic disparities and trends of food selections that would be masked by traditional survey-based estimation.

Keywords
Nutrition surveillance
Grocery transaction data
Ecological analysis
Community health assessment

Dietary habits and preferences are influenced by demographic, economic and socio-cultural factors, and built, policy and food marketing environments[1]. Neighbourhood heterogeneity of these attributes manifests as geographic disparities in diet quality and the burden of nutrition-related non-communicable diseases[1,2]. Geographic measurements of diets at a fine spatial scale (e.g. city subdivisions) are lacking but needed to identify at-risk communities, mobilise community advocacy for interventions tailored to local needs and evaluate geographically heterogeneous responses to public health policies[3–5].

Nutrition surveys administering dietary questionnaires are typically designed to measure population diets at the levels of 'subnational' geographic unit, such as a province or state. However, they are unable to estimate diets at smaller administrative areas due to limited sample sizes to attain statistically precise estimates[3,6,7]. The increasing demands for reliable statistics describing health status of

*Corresponding author: Email hiroshi.mamiya@mail.mcgill.ca

small subpopulations, coupled with the increasing non-response and cost of conducting many national surveys, calls for the integration of secondary data sources in population health assessment[7]. An emerging data source to capture community food selections is geocoded store-level grocery transaction records that are continuously generated by a geographically representative sample of retail food stores. We previously proposed a geographic indicator of the sales of soda through model-based smoothing of these data[8]. However, the estimated value of the indicator was not intuitively interpretable by practitioners and decision makers, as they represented the value of spatial random effect, which is the estimate of residual (unmeasured) effect, thus framed as 'residual' area-level demand of soda not accounted by area-level predictors.

In this research, we developed more interpretable community indicators of food purchasing by generating actual quantities of food purchased in each area through a probabilistic store catchment area analysis combined with store-level transaction data, which results in partitioning and allocation of store-level sales into surrounding areas proportional to store visit probability. The Huff gravity model generates a discrete (area-level) surface of the visit probability for each store as calculated from store size and travel distance from surrounding areas[9]. The model has been widely used to convert quantities of store-level sales into surrounding areas for the past four decades among retail management scientists[10,11].

Our primary objective was to develop small-area indicators of food purchasing from store-level sales within the Census Metropolitan Area (CMA) of Montreal using the Huff gravity model and to illustrate the use of these indicators in improving the estimation of area-level risk for type 2 diabetes mellitus (T2D) using a model-based small-area estimation of disease burden (disease mapping)[12].

The transaction data we obtained contained data for non-alcoholic beverages and solid food that consists of ultra-processed food (e.g. snacks) and three minimally processed food categories: vegetables, fruits and plain yogurt. As an initial demonstration for our methodology, we developed purchasing indicators for soda (carbonated soft drinks) and yogurt. Among sugar-sweetened beverages that are the largest source of added sugar in population diets, soda is one of the most important sources of total energetic intake in Canada and many economically developing and developed nations and has established associations with both T2D and obesity[13–16]. We also developed a separate indicator for diet soda (soda with artificial, low-calorie sweeteners substituting sugars). While its association with T2D is slightly weaker and less investigated than the non-diet soda[17], we were interested in inspecting its spatial distribution and association with T2D relative to that of soda.

Solid yogurt has been one of the fastest growing food groups in North America[18] with an important public health consequence: despite some evidence for health benefits[19], most yogurt products contain artificially added sugar[20], making yogurt an overlooked source of calories for Canadians, in particular among children who are facing an increasing incidence and regional disparity of T2D[21]. This makes yogurt a uniquely polarised food group, classified into two extremes of minimally processed (plain yogurt) and ultra-processed (flavoured) food categories[22]. For this reason, we were interested in developing separating indicators and contrasting the spatial distribution of the two yogurt categories, which are flavoured (sugar-added) yogurt and plain yogurt, the latter containing intrinsic (naturally derived from milk) sugars only. While the other minimally processed food items, fruits and vegetable, have an established or strong and negative association with many non-communicable diseases (most notably cancers and CVD), their associations with T2D – and the quality of evidence – is somewhat lower in comparison with yogurt[17,23], we therefore focused on the indicator for yogurt.

Methods

We conducted an ecological, cross-sectional study that developed area-level purchasing indicators for soda and yogurt food categories using store-level sales for the year 2012 within the Montreal CMA (delineation of the CMA is provided in Fig. 1), which contained a population of 3 824 211 in 2011[24]. The indicators were subsequently used to estimate spatially high-resolution T2D risk within the CMA for the same year.

Transaction data

Point-of-sales (i.e. store-level) transaction data are collected by several marketing firms that operate internationally. We purchased these data from Nielsen[25], which collects and centralises weekly transactions from a stratified random sample of geocoded chain retail food stores, with the strata defined by store size and store type, namely chain pharmacies, supermarkets and supercentres whose annual store-level revenue was greater than 2 million Canadian dollars. The store types are defined in online supplementary material, Supplemental Appendix S1 and Supplemental Table 1 in a separate file named as SupplementaryMaterial.docx.

There were 1097 chain food stores in the CMA of Montreal in 2012; 125 were sampled by the company to collect transaction data, and the remaining 972 chain stores were not sampled ('out-of-sample') stores, thus having unobserved or missing sales. The sampling frame included chain supermarket, pharmacy and supercentre chains; all of which sold soda and diet-soda products. Plain and flavoured yogurt were sold in a subset of 311 stores in Montreal (seventy-two sampled and 239 out-of-sample stores, consisting of all supermarket chains and one supercentre

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

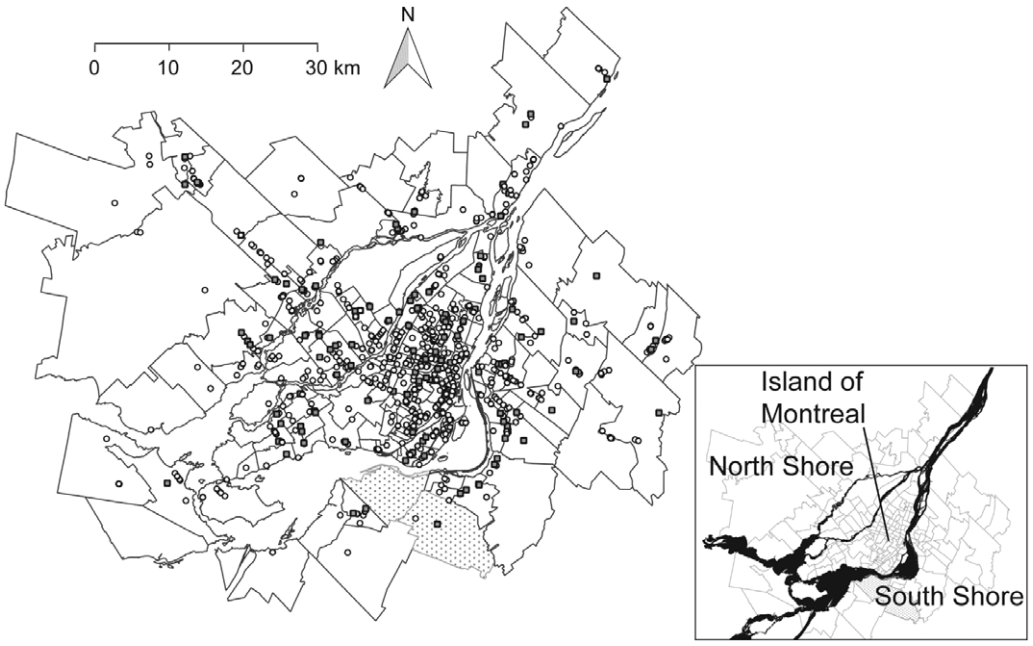

Fig. 1 Map of 193 neighbourhoods and two excluded neighbourhoods in the Census Metropolitan Area of Montreal, and an illustration of sampled and out-of-sample chain retail supermarkets, pharmacies and supercentres, 2012. Grey squares indicate sampled stores with sales data observed. Empty circles indicate out-of-sample stores, thus missing sales data. The location of points does not reflect exact location of stores, but randomly placed within the neighbourhood boundaries stores belong to. Two dotted areas are the federally registered First Nations communities without census data. These areas were excluded, leaving 193 neighbourhoods with white fill for this study. Black lines represent neighbourhood boundaries. Black fills in the inset map represent water. Census Metropolitan Area is defined as the aggregation of contiguous Canadian census subdivisions that are merged if extensive trips of residents occur. ■, Sampled store; ○, out-of-sample store; ⁘, Federally Registered First Nation Community (excluded area)

chain) as detailed in online supplementary material, Supplemental Fig. 1 in a separate file. An illustrative map of sampled and out-of-sample stores is shown in Fig. 1. The data are collected in many nations and shared through the company or affiliated academic institutions with the permission of the company. While we had access to transaction records generated by convenience stores, the accuracy of store locations with respect to our spatial unit of interest (neighbourhood, described below) was low for convenience stores, thus leading to misclassification of areas they belong to. We therefore did not include convenience stores in our analysis. In general, the geographic location of convenience stores appears to correlate with that of supermarkets (online supplementary material, Supplemental Fig. 2 a and b), likely reflecting the underlying population density; however, supermarkets had a slightly higher density in the centre of the island of Montreal.

Measure of store-level sales

We extracted transactions of food products belonging to the soda and yogurt categories. We did not have access to a comprehensive database recording nutritional composition of all food items sold in the study region, which would allow objective classification of food items by the presence of artificially added free sugar by manufacturers. We therefore separated diet soda and plain yogurt according to terms suggestive of these products (online supplementary material, Supplemental Appendix S2) and converted the quantity of sales into the Food and Drug Administration standardised serving (240 ml for beverages, 150 g for yogurt). Finally, we calculated the average store-level sales of the four food categories in 2012 by aggregating store-week sales of individual food items, see online supplementary material, Supplemental Appendix S3.

Spatial unit of analysis

Area-level food purchasing indicators were defined at the level of a small urban unit, Montreal neighbourhoods (193 areas, shown in Fig. 1), which were formed by merging contiguous Canadian census tracts to maintain the homogeneity of socio-demographic and economic attributes of residents within each neighbourhood[26]. The median number of residents per neighbourhood was 18 229 (interquartile range: 11 752–25 517).

Neighbourhood-level diabetes prevalence

As described previously[8,27], we defined T2D cases as 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 2-year period. The numbers of cases

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

and non-cases were determined from the provincial universal health insurance registry, the Régie de l'assurance maladie du Québec. After excluding individuals under the age of 2 or having gestational diabetes, prevalent cases and non-cases were linked to residential neighbourhoods within the CMA (n 193) by postal codes.

Statistical analysis

The analysis followed three steps. First, because most stores had missing sales (i.e. being out-of-sample stores), we predicted store-level sales for the four food categories among out-of-sample stores using a hierarchical Bayesian spatial model (Fig. 2a and b). In the second step, the predicted sales from stores were partitioned and allocated to surrounding neighbourhoods based on weights. The weights represent shopping flow, or visit probabilities, from each neighbourhood to food outlets (Fig. 2c) as calculated by the Huff gravity model. Each neighbourhood therefore is attributed a fraction of sales from surrounding stores proportional to the visit probabilities. In the third step, we illustrated the example utility of our small-area indicators as ecological covariates to improve the fit of a disease mapping model that estimates neighbourhood-level risk of T2D.

Step 1: Sales prediction model

The hierarchical Bayesian model to predict sales for the out-of-sample stores as a function of store- and area-level predictors of sales was described previously[8]. Specifically, the store-level sales of a given food category Q_{ij} at store j located in neighbourhood i in 2012 were natural log-transformed to follow a normal distribution. A subset of the sales vector Q_i had missing values to be predicted for out-of-sample stores and naturally treated as parameters to estimate following the Bayesian paradigm. The mean of Q_{ij}, denoted as μ_{ij}, was specified as

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

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

where the component $\phi_{c[j]}$ is a random effect representing store chain c to which store j belongs. The random effect accounts for chain-level environmental features, such as product pricing, availability (number of food items) and marketing activities such as flyers and display promotions, which tend to synchronise strongly across stores sharing the same chain identification codes.

Neighbourhood-level attributes are captured by Z_i. Its component A_i represents a vector of a store's neighbourhood socio-demographic and economic predictor of sales obtained from the 2011 National Household Survey[28]. They are area-level median family income, education as the proportion of individuals over the age of 25 who have post-graduate certificate or diploma, population density as the number of residents per square kilometre, 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. The corresponding coefficients are denoted as φ. An area-level random effect, S_i, accounted for a spatially correlated latent (residual) effect in sales as described in online supplementary material, Supplemental Appendix S4 along with the specifications of prior probabilities. Codes are provided in in a separate file named as SupplementaryTextFile1_code1.docx.

Step 2: Calculation of area-level purchasing from store-level sales using the Huff gravity model

The Huff model uses travel cost and store size to model the shopping trip of residents from their residential area to stores[9]. For a given food category, the probability, π_{ij}, of residents in neighbourhood i choosing each alternative store j is derived from a discrete store choice model:

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

for $i = \{1, \cdots, I\}$, where $I = 193$ neighbourhoods and $j = \{1, \cdots, J\}$, where $J = 1097$ stores for sodas and 311 stores for yogurts. The attraction of store, a_j, is frequently measured by store size (e.g. square footage) with the coefficient $\gamma 0$, which we approximated with the number of employees in each store. The variable d_{ij} represents the travel cost as defined by the shortest street distance in kilometres between the centroid of neighbourhood i and destination store j, with the distance decay coefficient $\lambda 0$. Our definition of distance between neighbourhoods and stores accounts for travel through the nearest bridge, if stores and neighbourhoods were separated by the body of water. We calculated the shortest network path for each store-neighbourhood pair using Dijkstra's routing algorithm available in the pgRouting module in the PostGIS geospatial database[29]. These store-visit probabilities are identical for the four food categories.

Empirically evaluated and commonly used value to accurately predict store visit in an urban setting is 2·0 for the distance decay (thus power decay of 2·0), and the value of 1·0 for the attraction coefficient[9,30–33]. While the coefficients are ideally estimated from a shopping-related local travel data[32,33], such data are uncommonly available to date (including our study) due to the cost of administering a large-scale mobility survey[33]. We therefore adopted the aforementioned values of $\lambda = 2.0$ and $\gamma = 1.0$. In sensitivity analyses, we varied the values of the attraction and distance decay parameters (online supplementary material, Supplemental Appendix S5).

The calculated store-visit probabilities were used to allocate store-level sales to surrounding neighbourhoods. For $I = 193$ neighbourhoods and $J = 1097$ (311 stores for yogurts) stores, the visit probability π_{ij} populates $I \times J$

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

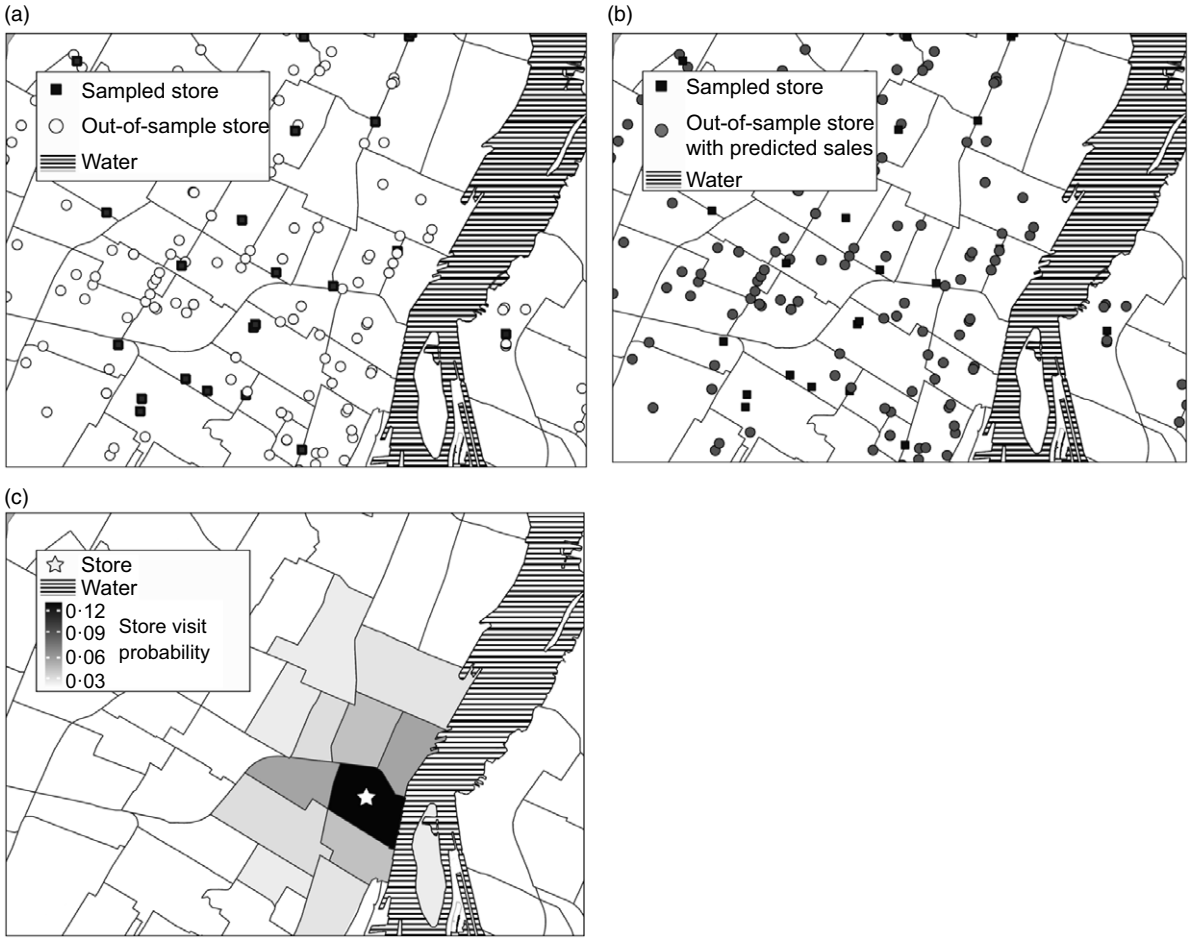

Fig. 2 Illustration of analytical approach to generate area-level indicator of purchasing from store-level sales data. Panel (a) illustrates hypothetical geographic location of sampled (squares with black fill) and out-of-sample stores (empty circles), the latter missing sales records. Panel (b) illustrates out-of-sample stores with predicted distribution of sales (dark fill in circles) as well as the sampled stores with observed sales. Panel (c) illustrates the neighbourhood-level store visit probabilities for a hypothetical store (star symbol) as generated by the Huff gravity model, which represent a probabilistic catchment zone of the store in the form of discrete (area-level) surface of store visit probabilities. The product of these area-level visit probabilities and population size in neighbourhood represents weights to split and allocate predictive distribution of sales (for out-of-sample stores) or observed sales (for sampled stores) into surrounding areas, with the quantity of store-level sales fixed. Note that there are as many probability maps as the number of other stores that are not displayed in this map. Solid lines represent boundaries of Montreal neighbourhood

origin-destination (i.e. neighbourhood-store) matrix, whose rows are multiplied with population size of area i, C_i:

$$G = \begin{bmatrix} \pi_{11}c_1 & \pi_{12}c_1 & \pi_{13}c_1 & & \pi_{1J}c_1 \\ \pi_{21}c_2 & \pi_{22}c_2 & \pi_{23}c_2 & & \pi_{2J}c_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_{I1}c_I & \pi_{I2}c_I & \pi_{I3}c_I & & \pi_{IJ}c_I \end{bmatrix},$$

where the measurement of C was provided by the 2011 Canadian Census[28]. The column-normalised version of the matrix, G^* (i.e. columns sum to 1, a constraint used to ensure that the store-level quantity of sales was not changed), is multiplied with the $J \times 1$ vector of the exponentiated predictive distribution of store-level sales Q to generate the vector of area-level purchasing quantity in serving, X as

$$X = G^* exp(Q).$$

Thus, $X_i = \sum_{j=1}^{J} \pi_{ij} C_i exp(Q_j)$. In other words, X_i represents the predicted distribution of purchasing quantity for neighbourhood i calculated as the weighted sum of the posterior distribution of the vector Q consisting of predicted sales from J stores, for a given food category. We then divided these quantities by the number of residents in each neighbourhood to generate our indicator of small-area purchasing that represents *per-person* purchasing quantity in each neighbourhood.

We note that there were fourteen inhabited small areas (fourteen rows in G matrix above) that contained no supermarket, pharmacy and supercentre in the CMA of Montreal in 2012, even though residents in these areas shop at stores in surrounding areas. The gravity model

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

allowed generating the measure of food purchasing in these areas despite not containing any stores; in other words, the computed sum of these fourteen rows across columns in the matrix is non-zero. This is because these areas received a portion of sales from each store, or $\pi_{ij}C_i exp(Q_j)$, from other neighbourhoods proportional to the value of store visit probability and population size in these areas. These values reflect the shopping-related flow of residents across neighbourhood boundaries. Note that without consideration of mobility across neighbourhoods, the store visit probability will be binary, simplifying to $\pi_{ij} = 1$ if stores j is located in neighbourhood i, and zero otherwise. Therefore, the matrix will have a large number of zeros, unrealistically assuming that residents shop only in store(s) located in their own neighbourhood.

Step 3: Application of the indicators to a model-based small-area estimation of diabetes risk

We added the small-area indicators of purchasing into disease mapping, the primary spatial epidemiologic and geographic surveillance method to estimate areal-level disease risk using a hierarchical Bayesian model[34]. The above-mentioned prevalent T2D count in the i^{th} area, y_i, is assumed to follow a Poisson distribution with mean $e_k\theta_k$, where e_k is the expected number of cases (i.e. offset) calculated by indirect standardisation with age as detailed in online supplementary material, Supplemental Appendix S6. The outcome of interest for disease mapping, the area-specific relative risk, θ_i represents the deviation of risk from the expected count: an area i has higher occurrence of cases if the posterior summary of θ_i (e.g. 95 % credible interval (CI)) is greater than 1. The relative risk is modelled as:

$$log(\theta_i) = \beta_0 + G_i\gamma + X_i\beta + b_i,$$

where β_0 is an intercept, and G_i is a vector of area-level covariates associated with T2D with the corresponding coefficient vector γ. These covariates are education and median family income used in the sale description models and the proportion of immigrants calculated from the 2011 Canadian National Household Survey[35]. The covariates also include the availability of recreational facilities encouraging physical exercise (the number of facilities per resident) obtained from the Canada Business Point data, which are annually updated business enumeration data and validated previously for the accuracy of business locations[36,37]. The availability of recreational facilities was calculated as the number of facilities divided by 1000 residents for each neighbourhood, where recreational facilities were defined based on the Standard Industry Classification codes as previously described[38]. The vector of our indicators (neighbourhood-level per-resident purchasing quantity for the four food categories) and their coefficients are X_i and β, respectively. To account for spatial autocorrelation and Poisson overdispersion of the disease risk, we added an area-level random effect b_i as specified in online supplementary material, Supplemental Appendix S7; codes are also provided (SupplementaryTextFile2_code2.docx).

To investigate whether the addition of our indicators improved the disease risk estimation or not, we investigated their posterior 95 % CI. To supplement model selection where we compared goodness of model fit with and without the indicators, we computed fully Bayesian estimates of pointwise predictive density by Watanabe–Akaike information criterion and approximate leave-one-out cross-validation[39,40]. A lower value indicates better model fit. However, evaluating goodness of it using these metrics requires some caution, as the measures of pointwise predictive errors generally do not acknowledge the spatially structured nature of data[39,40]. We also note that the fully Bayesian approach would jointly estimate the posterior distribution of T2D risk (step 3) and indicators (step 1 and 2), therefore propagating the uncertainty of step 1 and 2 to the distribution of the parameters generated in step 3. We did not take this approach and ran step 3 independently, since the indicator generation step, the main goal of this research and downstream applications of these indicators for public health research and practice (e.g. surveillance) will be inherently disjointed. This is because the indicators are likely to be a priori estimated and disseminated by analysts for use by third parties.

We note that both sales prediction and T2D model include income and education as their covariate. Thus, if the store-level sales were strongly or predominantly driven by income and education, our proposed indicators may simply represent the information (i.e. spatial distribution and posterior distribution of regression coefficients) of these two variables, which is redundant and may in fact exhibit collinearity with neighbourhood education and income in the T2D model. Therefore, we performed another sensitivity analysis, where we generated the proposed indicators with modified sales models excluding the neighbourhood-level covariates, including education and income. The estimated posterior mean and 95 % CI of the regression coefficients for income, education and the proposed indicators in the T2D model were then compared with that of the coefficients in the T2D model in the main analysis.

Results

Descriptive analysis

Table 1 summarises the store-level quantity of standardised servings sold for each food category, which shows a wide variation of sales reflecting store size and chain type. The mean and median quantities of soda and flavoured yogurt servings sold were considerably larger than their low-calorie (diet) counterparts. Neighbourhoods of sampled stores were representative of all neighbourhoods in the Montreal CMA with respect to the distribution of

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

Table 1 Characteristics of store-level transactions by food category for sampled stores in the Census Metropolitan Area of Montreal, 2012*

	Min	Mean	Median	IQR	Max
Soda†	8·4	18 510·3	13 071·6	1322·7–26 957·0	108 771·0
Diet soda†	2·6	6939·4	5214·5	664–11 588·9	29 466·2
Flavoured yogurt‡	1679·4	6809·6	6454·6	4479·4–8711·0	17 361·3
Plain yogurt‡	81·3	698·9	588·2	333·4–907·3	2556·3

IQR, interquartile range; Max, maximum; Min, minimum.
*These quantities are the average of weekly beverage sales by standard serving.
†Sales of soda represent transactions in supermarkets, supercentres, and pharmacies.
‡Sakes if yogurts represent transaction in supermarkets and one supercentre chain.

Table 2 Characteristics of neighbourhoods on which sampled chain stores were located (eighty-three areas) and all neighbourhoods (193 areas) in the Census Metropolitan Area of Montreal, 2011 Canadian National Household Survey

	Neighbourhoods with sampled stores (83 areas)				All neighbourhoods (193 areas)			
Neighbourhood characteristics	Min	Median	IQR	Max	Min	Median	IQR	Max
Count of T2D	53	385	300–483	1617	43·0	319·0	211·0–436·0	1617·0
Prevalence of T2D (%)	3·3	6·9	6·1–8·3	10·4	3·1	6·8	5·8–8·2	10·7
Education*	29·6	67·6	61·5–74·8	90·3	29·6	66·3	60·5–74·8	90·3
Median family income (in 10 000 Canadian dollars)	2·0	7·0	5·8–8·7	16·7	2·0	6·9	5·5–8·5	16·7
% of immigrant	1·6	20·4	8·1–32·6	63·1	1·6	20·9	9·5–32·7	63·6
% of residents under 18 years old	2·4	20·5	17·2–23·3	32·9	2·4	20·9	17·5–23·3	32·9
% of employed residents among labour force	18·5	59·0	54·1–66·9	76·8	18·5	59·0	54·1–66·3	76·8
Average family size	1·1	3·0	2·9–3·1	3·5	1·1	3 .0	2·8–3·1·0	3·5
Population density (residents per square kilometre)	109·2	2298·6	1074·1–5345·1	19 606·1	54·7	3348	1093·5–5744·8	19 606·1
Number of recreational facilities per 1000 residents†	0·1	0·4	0.–0·5	1·3	0·0	0·3	0·2–0·5	2·8

IQR, interquartile range; Max, maximum; Min, minimum; T2D, type 2 diabetes mellitus.
*% of residents with post-graduate diploma or certificate among ages greater than 25.
†The number of recreational facilities was obtained from Canada Business Point data (see main text).

neighbourhood characteristics, except population density (Table 2).

Sales were predicted reasonably accurately: fitted and observed sales of soda and diet soda appeared to correlate well (Fig. 3a and b), although fitted sales of flavoured and plain yogurt (Fig. 3c and d) slightly underestimated sales at larger stores and overestimated sales at smaller stores. Store chain random effects had a prominent association with sales, especially for soda and diet soda in pharmacy chains (Fig. 4), whereas the fixed effect of standardised area-level covariates showed only modest associations with sales, except education (positive association) and income (negative) for the sales of plain yogurt.

Figure 5 maps the estimated posterior means of our indicators representing neighbourhood-level average weekly purchasing per resident for soda, diet soda, flavoured yogurt and plain yogurt. The spatial trend for plain yogurt was distinct: higher quantities of purchasing were clustered in the centre of the island of Montreal, while for the other food categories, purchasing was more intense in the north of the island and in the North- and Southshore. The corresponding posterior standard deviation of the indicators (Fig. 6) did not show clear spatial trends, but tended to be larger in neighbourhoods with a smaller number of stores (online supplementary material, Supplemental Fig. S3). The sensitivity analyses varying the distance decay coefficient affected spatial smoothing for area-level purchasing of all food categories (online supplementary material, Supplemental Figs. 4–7, b and c), while this smoothing was not prominent for the sensitivity analysis using the smallest value of the attraction coefficient (online supplementary material, Supplemental Figs. 4–7, d). Across all sensitivity analyses, the indicators appear to consistently capture large-scale spatial trends in purchasing, that is, clustering of purchasing quantities in the Island of Montreal for plain yogurt, and the opposite pattern for the other food categories.

Table 3 shows the posterior mean and 95 % CI of the neighbourhood-level T2D risk associated with neighbourhood-level covariates. The indicator for plain yogurt had a negative association with the relative risk of T2D (posterior mean: 0·93, 95 % CI 0·89, 0·96), while the indicator for flavoured yogurt showed a positive association (1·08, 95 % CI 1·02, 1·14). In contrast, soda did not show a conclusive association (95 % CI 0·97, 1·05). The indicator of diet soda was removed from the model due to its strong correlation with the soda indicator. The fit of the model including our indicators was superior to the model without the indicators

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

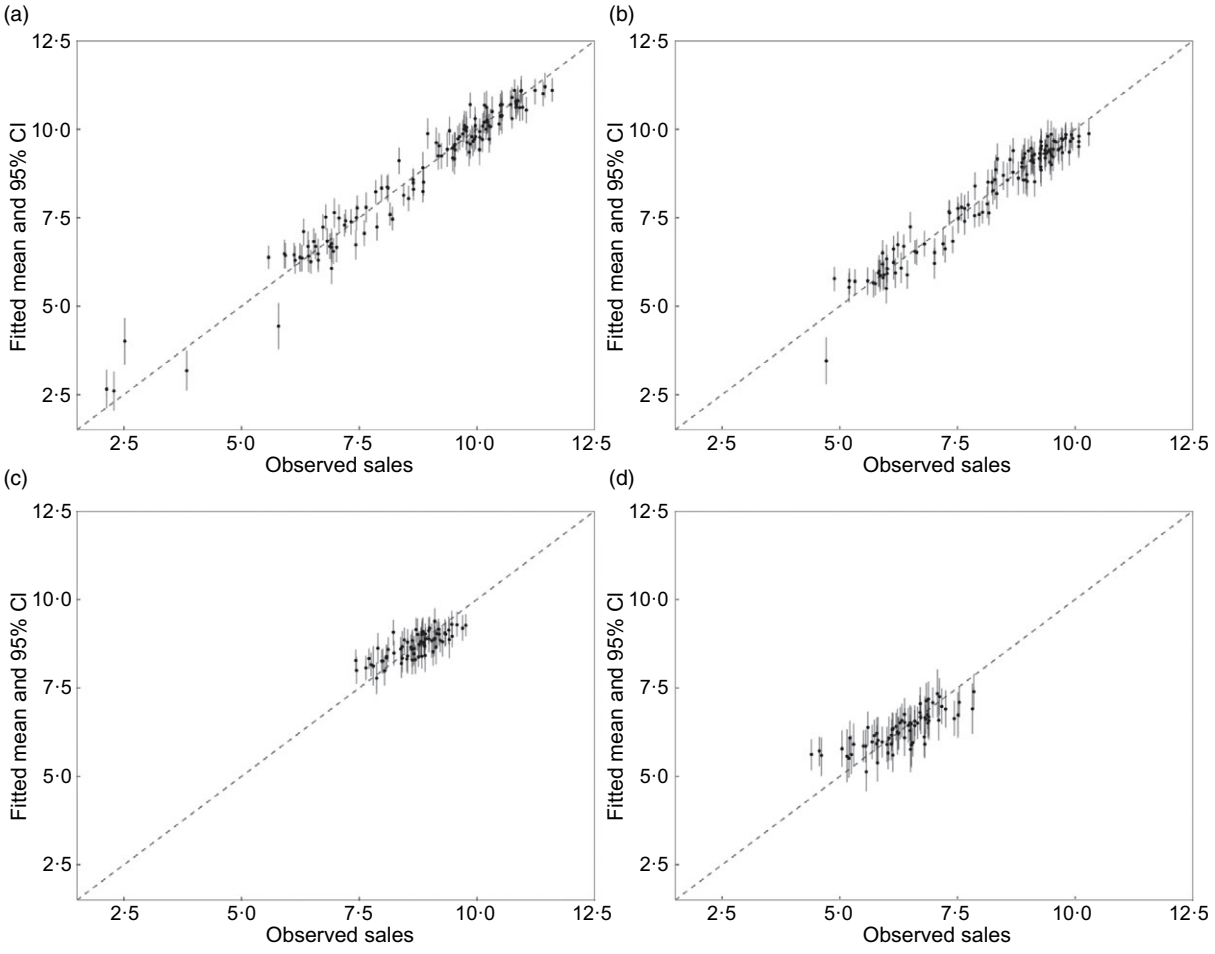

Fig. 3 Fitted and observed log sales of (a) soda, (b) diet soda, (c) flavoured yogurt and (d) plain yogurt in the sales model. Vertical solid line indicates 95 % posterior credible interval of fitted log sales. Dashed line is reference slope, where y = x

(leave-one-out cross-validation = 1874 with the indicators *v.* 1885 without; Watanabe–Akaike information criterion = 1765 with the indicators *v.* 1772 without), therefore suggesting a higher accuracy of estimated small-area risk of T2D upon addition of the indicators.

Our sensitivity analysis generating the proposed indicators from varying value of the distance decay and store attraction coefficients also resulted in superior model fit over the model without the indicators (online supplementary material, Supplemental Table 3). However, the association of the indicator for the yogurt categories tended to be attenuated by the greater spatial smoothing of purchasing induced by the lower value of the distance decay coefficient in the Huff model (online supplementary material, Supplemental Fig. 8). Another sensitivity analysis, removing the neighbourhood socio-demographic and economic predictors from the sales models and fitting the T2D models to the resulting purchasing indicators, showed largely similar posterior mean of the purchasing indicators to those generated in the main analysis (online supplementary material, Supplemental Fig. 9 a–c), except for the indicator for plain yogurt that showed a noticeable difference from the main analysis for neighbourhoods with higher values of purchasing quantity (online supplementary material, Supplemental Fig. 9d). The finding led to unsurprising results in the downstream application of these indicators to the T2D model; the posterior summary of the coefficient for income, education and our indicators in the T2D models was very similar to that of coefficients in the main analysis (online supplementary material, Supplemental Table 4).

Discussions

We generated small-area public health indicators of food purchasing from routinely generated store-level grocery sales data to addresses the current lack of measurements to identify neighbourhood heterogeneity and patterning of food selections.

Visually similar spatial distributions of the indicators for soda and diet soda purchasing may reflect non-differential preference for diet soda products. However, the spatial

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

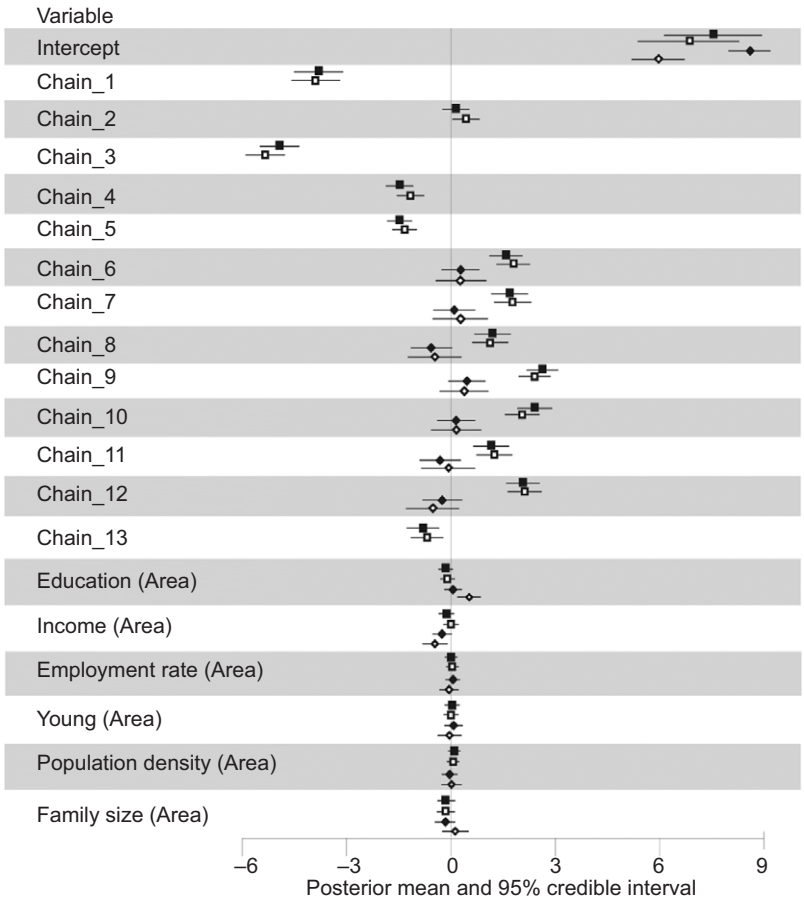

Fig. 4 Posterior summary of sales model for each food category. Models for each food category were ran separately. The values of the coefficients represent association with log store-level sales in servings. Store chain indicator represents a retail chain identifier for random effect; unlike a dummy variable in a fixed effect model, there is no baseline category to which the indicator of store chain is compared. Chains 1–5 are pharmacies, chains 6–11 are supermarkets and chains 12 and 13 are supercentres. Fixed effects representing neighbourhood-level predictor of sales were mean cantered and scaled at one standard deviation. The covariate 'young' represents the proportion of residents under 18 years old, and the covariate 'family size' represents the mean number of family members. The sales models for flavoured and plain yogurts did not include sales in chains 1–5 and 13, as pharmacies and one supercentre chain rarely sold yogurt. ■, Soda; □, diet soda; ◆, flavoured yogurt; ◇, plain yogurt

trend of purchasing for plain yogurt was distinct from that of sodas and flavoured (sugar-added) yogurt. Given that store chain random effects in the sales prediction model had a considerably stronger association with sales than neighbourhood attributes such as income and education, the geographic patterns of our purchasing indicators (and therefore community food selection) maybe predominantly driven by the neighbourhood composition of store chains. Because plain yogurt was the only food category whose store-level sales were conclusively (albeit modestly) associated with neighbourhood education and income, it is not surprising that only the indicator of this category resulted in a notable disagreement with the main analysis when the neighbourhood socio-economic attributes from the sales prediction model were removed. However, the largely invariant association of T2D risk with neighbourhood income and education as well as the proposed indicators in this sensitivity analysis suggests that our indicators contain information unique to the neighbourhood socio-economic determinants, rather than simply being their correlates.

The quantity of flavoured yogurt purchased was far higher than that of plain yogurt in all neighbourhoods, a finding that may support yogurt being ranked in the top ten sources of total sugar intake among Canadian children[21]. Our indicator for plain yogurt was associated with a decreased areal risk of T2D, while the association was positive for flavoured yogurt. While accumulating evidence suggests the association of frequent yogurt intake with a decreased risk of non-communicable diseases including T2D[19], the potential benefit may be limited to plain yogurt, if the consumption of flavoured yogurt contributes to excess energetic intake that is linked to obesity and T2D[20]. Our study is ecological and thus does not permit an individual-level epidemiologic link between yogurt and non-communicable diseases. However, the distinct

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

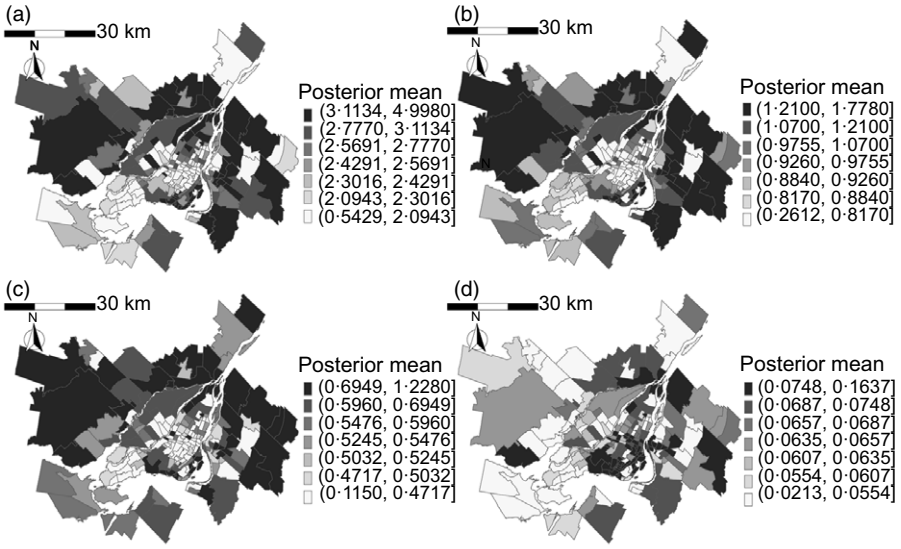

Fig. 5 Posterior mean of neighbourhood indicator for (a) soda, (b) diet soda, (c) flavoured yogurt and (d) plain yogurt in the Census Metropolitan Area of Montreal, 2012. The greyscale key to the right of each map indicates posterior mean of purchasing quantity per resident in serving. Note that the quantities in the greyscale keys are not standardised across food categories, as the quantities were very different in scale

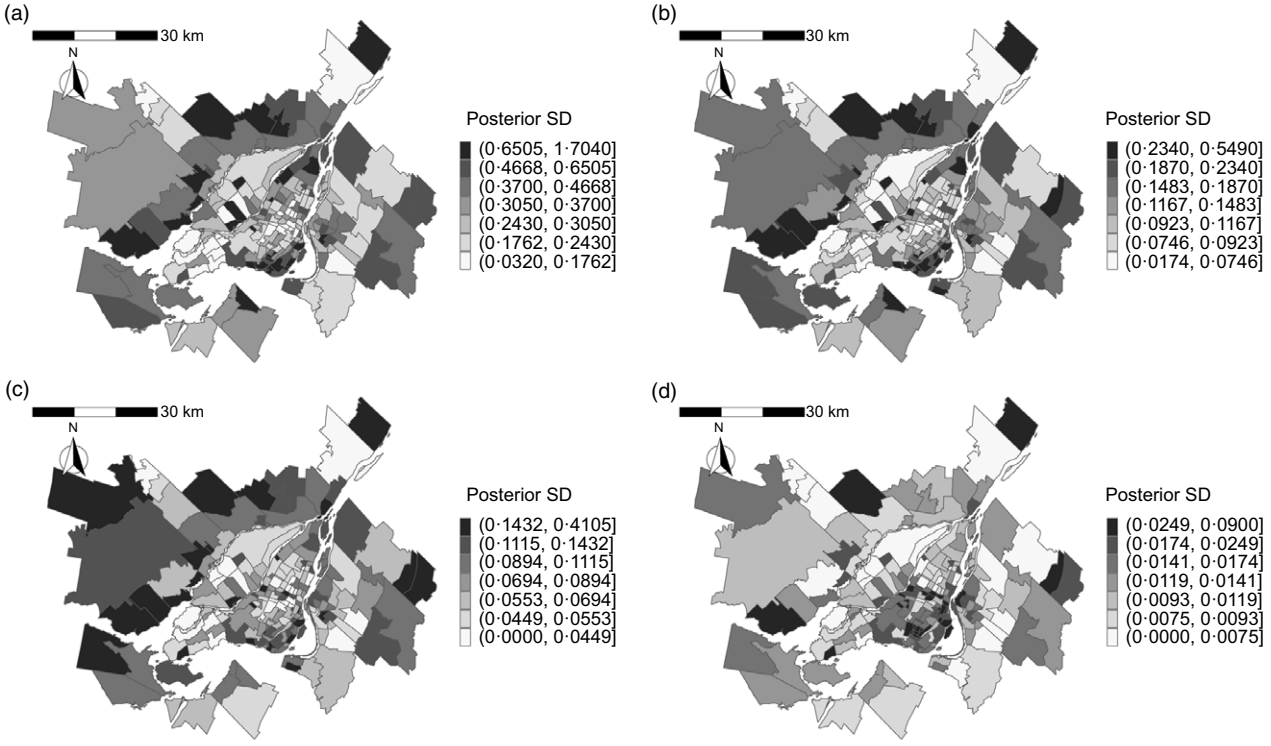

Fig. 6 Posterior standard deviation of neighbourhood indicator for (a) soda, (b) diet soda, (c) flavoured yogurt and (d) plain yogurt in the Census Metropolitan Area of Montreal, 2012. The greyscale keys to the right of each map indicate posterior standard deviation of purchasing quantity per resident in each neighbourhood. Note that the quantities in the greyscale keys are not standardised across food categories, as the quantities were very different in scale

spatial distribution and the direction of the association with T2D across the yogurt categories indicate the importance of monitoring yogurt items (and likely other food products) separately based on sugar composition.

Ongoing measurement of neighbourhood food selection through nutrition surveys is infeasible given the lack of adequate sample sizes in each area, and often many areas have zero survey participants as the spatial unit of

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

Table 3 Posterior mean and 95 % credible interval of exponentiated coefficients and model fit of neighbourhood-level (*n* 193) diabetes risk model, Census Metropolitan Area of Montreal, 2012*

	Without purchasing indicators		With purchasing indicators	
Parameter	Mean	95 % CI	Mean	95 % CI
Intercept	1·10	1·03, 1·17	1·14	1·06, 1·22
Neighbourhood-level attributes				
Education†	0·92	0·89, 0·95	0·96	0·92, 1·00
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
Neighbourhood-level purchasing indicators				
Soda†,‡			1·01	0·97, 1·05
Yogurt (plain)†,‡			0·93	0·89, 0·96
Yogurt (flavoured) [a, b]			1·08	1·02, 1·14
Model fit				
LOO	1885		1874	
WAIC	1772		1765	

95 % CI, 95 % credible interval; LOO, leave-one-out cross-validation; mean, posterior mean; recreation, recreational facility per resident; WAIC, Watanabe–Akaike information criterion.
*The indicator of diet soda was removed from the model due to its strong correlation with soda indicator.
†Variables were mean cantered and scaled to one standard deviation, and the regression coefficients were exponentiated. The value of coefficients represents neighbourhood-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.
‡The value of the indicators represents neighbourhood-level purchasing quantity per resident.

analysis becomes smaller in physical size. A potential remedy, increasing sample size by combining multiple years of data collection, is not possible for nutrition surveys which are infrequently conducted. Our study overcomes this sparseness of survey sampling using the predicted store-level aggregated sales combined with gravity model that allocated sales from stores to neighbouring areas. Alternative and previously utilised data source for small-area estimation is large sample loyalty card transaction data that are generated by millions of members; however, the representativeness of such data is limited to the participating members of loyalty club in a single major supermarket chain[41,42], potentially excluding subpopulations under an elevated risk of obesity who may preferentially use discount supermarkets[43]. To our knowledge, ours is the first small-area estimation of food selection that can be generated at ongoing basis and reflects purchasing behaviours from multiple chains of supermarkets and other store types. Equally importantly, many studies analysing geographically indexed data fail to address spatial autocorrelation of measurements[44], which can overestimate precision and may bias the association of interest. Our sales and T2D risk model accounted for latent spatial effects due to spatially structured unmeasured variables.

An important limitation of our work is the exclusion of convenience stores, a potentially important source of *soda* sales (yogurts were rarely sold in convenience stores). Supermarkets are reported to be the dominant location of unhealthy food purchasing[45], and our previous study indicates chain supermarkets sold considerably higher quantity of soda than convenience stores (median sales: 22 026 *v.* 898 servings, respectively)[46]. However, convenience stores outnumber supermarket: there were 2732 chain and independent convenience stores and 662 supermarkets in the Metropolitan Montreal in 2012. It is thus possible that the estimated values of the indicators were biased and led to the inconclusive association of the indicator of soda and diet soda with the risk of T2D. While the current analysis better captures food categories not sold in convenience stores, further research and data to obtain accurately geocoded sales of these store are needed. As well, purchasing data capture food selection rather than consumption, although sugar intake measured by household purchasing and consumption data from recall appears to correlate well[47]. We also note that our approach to classify food items into sugary and non-sugar (absence of artificially added free sugars) may not be nutritionally objective, since the accuracy of product descriptor to categorise food items in transaction data is not known and some key terms are ambiguous, reflecting nutritional claims by food industries rather than actual sugar content. As an example, food items whose descriptor contains product claim label such as 'no added sugar' may still contain sugar, albeit less energy dense than items without these claims, depending on national labelling regulation[48]. Finally, while the law of retail gravitation suggests that travel distance and store size are the key factors influencing store selection, other potential predictors of store visit, such as pricing (discount chain or not) and in-store marketing activities, should also be added to the gravity model in future applications[49,50].

Future research includes scaling-up of food categories beyond soda and yogurt to capture comprehensive purchasing patterns that will better inform programmes to improve diets and estimate disease risk. While we focused on spatial analysis and used disease mapping to illuminate the application of our indicators, our work leads to a window of opportunities to enhance population assessment of food selections, including the incorporation of the temporal dimension to generate weekly or monthly evolution and fluctuation of neighbourhood-level purchasing patterns in response to socio-economical events and interventions. Such work will complement relatively infrequent updates of population health assessment driven by national nutrition surveys, which is, in case of Canada, conducted once every 10 years[51]. As well, the increasing availability of spatially detailed travel data from cell phone records, with appropriate anonymisation of travellers, implies new research opportunities to estimate the distance decay and attraction parameter based on mobility patterns specific to the study population[52,53]. Such data will also allow researchers to learn distance decay and attraction coefficients specific to store types, as we expect that stores utilised for smaller and shorter shopping excursions, such as pharmacies, have a larger distance decay than

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press

supermarkets[31]. As previously demonstrated, empirical mobility data may also allow learning the variation of distance decay coefficient across areas (e.g. areas with low median household income may show larger distance decay due to lack of access to vehicle for shopping)[54].

Given the increasingly acknowledged local disparities of chronic diseases burdens and neighbourhood effects of health, measurement capacity of public health surveillance and research should encompass small-area heterogeneity of behavioural risk factors including food purchasing[7,55,56]. Our analysis applied to ubiquitous grocery sales data provides a foundation to expand the measurement capacity.

Acknowledgements

Acknowledgements: None. *Financial support*: The results reported herein correspond to specific aims of grant 1516-HQ-000069 to investigator David Buckeridge from the Public Health Agency of Canada. This work was also supported by an Institut de valorisation des données (post-doctoral fellowship awarded to Hiroshi Mamiya), the Canadian Institute for Health Research (Applied Public Health Chairs program, grant number: CPP-137904 awarded to David Buckeridge and Foundation Grant, grant number: CIHR FDN-167267 awarded to Erica Moodie). The funders had no role in the design, analysis and writing of this article. *Conflict of interest*: None. *Authorship*: The study was conceived and designed by H.M. AM.S. and E.E.M.M. provided inputs on the statistical analysis and interpretation of the results. Y.M. and D.L.B. provided the data and substantive knowledge to aid study design and interpretation of the results. Data analysis and drafting of manuscript was performed by H.M. All authors reviewed and commented on the manuscript, and they approved the final manuscript. *Ethics of human subject participation:* This study used secondary data that are aggregated store-level measurements of consumer purchasing, rather than individual consumer-level data. The study therefore did not require a separate written or verbal consent from human subjects.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1368980021003645

References

1. Swinburn BA, Sacks G, Hall KD *et al.* (2011) The global obesity pandemic: shaped by global drivers and local environments. *Lancet* **378**, 804–814.
2. Dekker LH, Rijnks RH, Strijker D *et al.* (2017) 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* **14**, 1–15.
3. Cheadle A, Sterling TD, Schmid TL *et al.* (2000) Promising community-level indicators for evaluating cardiovascular health-promotion programs. *Health Educ Res* **15**, 109–116.
4. Barnhill A, Palmer A, Weston CM *et al.* (2018) Grappling with complex food systems to reduce obesity: a US public health challenge. *Public Health Rep* **133**, 44S–53S.
5. Bauer UE, Briss PA, Goodman RA *et al.* (2014) Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *Lancet* **384**, 45–52.
6. Seliske L, Norwood TA, McLaughlin JR *et al.* (2016) Estimating micro area behavioural risk factor prevalence from large population-based surveys: a full Bayesian approach. *BMC Public Health* **16**, 478.
7. National Academies of Sciences, Engineering, and Medicine. Statistical Methods for Combining Multiple Data Sources (2017) *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*. Washington, DC: The National Academies Press.
8. Mamiya H, Schmidt AM, Moodie EEM *et al.* (2019) 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* **188**(9), 1713–1722.
9. Huff DL (1964) Defining and estimating a trading area. *J Mark* **28**, 34–38.
10. Joseph L & Kuby M (2011) Gravity modeling and its impacts on location analysis. In *Foundations of Location Analysis*, pp. 423–443 [Eiselt HA & Marianov V, editors]. Boston, MA: Springer US.
11. Suhara Y, Bahrami M, Bozkaya B *et al.* (2019) Validating gravity-based market share models using large-scale transactional data. http://arxiv.org/abs/1902.03488 (accessed 2019 May 5).
12. Wakefield JC, Best NG & Waller L (2001) Bayesian approaches to disease mapping. In *Spatial Epidemiology: Methods and Applications*. Oxford University Press; https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198515326.001.0001/acprof-9780198515326-chapter-7 (accessed 2019 August 26).
13. Hu FB (2013) 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* **14**, 606–619.
14. Langlois K, Garriguet D, Gonzalez A *et al.* (2019) Change in total sugars consumption among Canadian children and adults. *Health Rep* **30**: 10–19.
15. Popkin BM & Hawkes C (2016) The sweetening of the global diet, particularly beverages: patterns, trends and policy responses for diabetes prevention. *Lancet Diabetes Endocrinol* **4**, 174–186.
16. Malik VS, Schulze MB & Hu FB (2006) Intake of sugar-sweetened beverages and weight gain: a systematic review. *Am J Clin Nutr* **84**, 274–288.
17. Neuenschwander M, Ballon A, Weber KS *et al.* (2019) Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. *BMJ* **366**, l2368.
18. Fernando J (2013) Yogurt Market: Current Status and Consumption Trends. https://open.alberta.ca/dataset/b5d936eb-2127-424e-b1b8-818c486d12aa/resource/1de9e2f1-e17f-4ae2-a1a8-65eb458b44f1/download/jeewaniyogurtmarketrevisedjune-112014.pdf.
19. Marette A & Picard-Deland E (2014) Yogurt consumption and impact on health: focus on children and cardiometabolic risk. *Am J Clin Nutr* **99**, 1243S–1247S.
20. Moore JB, Horti A & Fielding BA (2018) Evaluation of the nutrient content of yogurts: a comprehensive survey of

yogurt products in the major UK supermarkets. *BMJ Open* **8**, e021387.
21. Statistics Canada (2019) Change in Total Sugars Consumption among Canadian Children and Adults. https://www150.statcan.gc.ca/n1/pub/82-003-x/2019001/article/00002-eng.htm (Accessed 2019 July 21).
22. Monteiro C, Cannon G, Lawrence M *et al.* (2019) *Ultra-Processed Foods, Diet Quality, and Health Using the NOVA Classification System*. Rome: FAO.
23. Wallace TC, Bailey RL, Blumberg JB *et al.* (2020) Fruits, vegetables, and health: a comprehensive narrative, umbrella review of the science and recommendations for enhanced public policy to improve intake. *Crit Rev Food Sci Nutr* **60**, 2174–2211.
24. Government of Canada, Statistics Canada (2012) Statistics Canada: 2011 Census Profile. 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 (accessed 2015 December 29).
25. Nielsen. (2020) Retail Measurement Data. https://www.nielsen.com/us/en/solutions/measurement/retail-measurement (accessed 2021 February 26).
26. Ross NA, Tremblay S & Graham K (2004) Neighbourhood influences on health in Montréal, Canada. *Soc Sci Med* **59**, 1485–1494.
27. Clottey C, Mo F, LeBrun B *et al.* (2001) The development of the National Diabetes Surveillance System (NDSS) in Canada. *Chronic Dis Can* **22**, 67–69.
28. Government of Canada, Statistics Canada. (2013) 2011 National Household Survey Profile – Census Metropolitan Area/Census Agglomeration. 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 (accessed 2016 May 16).
29. PgRouting development team. pgRouting Project — Open Source Routing Library. https://pgrouting.org/ (accessed 2021 May 23).
30. Dolega L, Pavlis M & Singleton A (2016) Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. *J Retail Consum Serv* **28**, 78–90.
31. Sandhu DS (1989) *Comparison of Aggregate and Disaggregate Models in Predicting Shopping Centre Patronage [Thesis]*. Simon Fraser University. http://summit.sfu.ca/item/4792 (accessed 2020 May 28).
32. Nakanishi M & Cooper LG (1974) Parameter estimation for a multiplicative competitive interaction model: least squares approach. *J Mark Res* **11**, 303–311.
33. Huff DD & McCallum BM (2008) Calibrating the Huff model using arcGIS Business Analyst. An ESRI White Paper. https://www.google.com/search?client=ubuntu&channel=fs&q=Calibrating+the+Huff+Model+Using+ArcGIS+Business+Analyst.&ie=utf-8&oe=utf-8 (accessed 2020 October 5).
34. Waller LA & Carlin BP (2010) Disease mapping. *Chapman Hall CRC Handb Mod Stat Methods* **2010**, 217–243.
35. Statistics Canada.(2011) National Household Survey User Guide. https://www12.statcan.gc.ca/nhs-enm/2011/ref/nhs-enm_guide/index-eng.cfm (accessed 2020 January 20).
36. Daepp MI & Black J (2017) Assessing the validity of commercial and municipal food environment data sets in Vancouver, Canada. *Public Health Nutr* **20**, 2649–2659.
37. Pitney Bowes Canada (2012) *Product Documentation: Canada Business Data*. Pitney Bowes Canada. https://www.pitneybowes.com/ca/en (accessed January 2021).
38. Boone-Heinonen J, Diez-Roux AV, Goff DC *et al.* (2013) The neighborhood energy balance equation: does neighborhood food retail environment + physical activity environment = obesity? The CARDIA study. *PLOS ONE* **8**, e85141.
39. Vehtari A, Gelman A & Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* **27**, 1413–1432.
40. Gelman A, Hwang J & Vehtari A (2014) Understanding predictive information criteria for Bayesian models. *Stat Comput* **24**, 997–1016.
41. Howard Wilsher S, Harrison F, Yamoah F *et al.* (2016) 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* **13**, 21.
42. Aiello LM, Schifanella R, Quercia D *et al.* (2019) Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Sci* **8**, 14.
43. Chaix B, Bean K, Daniel M *et al.* (2012) Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). *PLoS ONE* **7**, e32908.
44. Lamb KE, Thornton LE, Cerin E *et al.* (2015) Statistical approaches used to assess the equity of access to food outlets: a systematic review. *AIMS Public Health* **2**, 358–401.
45. Vaughan CA, Cohen DA, Ghosh-Dastidar M *et al.* (2016) Where do food desert residents buy most of their junk food? Supermarkets. *Public Health Nutr* **20**, 2608–2016.
46. Mamiya H, Moodie EEM, Ma Y *et al.* (2018) 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. *Int J Epidemiol* **47**(6), 1877–1886.
47. Appelhans BM, French SA, Tangney CC *et al.* (2017) To what extent do food purchases reflect shoppers' diet quality and nutrient intake? *Int J Behav Nutr Phys Act* **14**. doi: 10.1186/s12966-017-0502-2.
48. Bernstein JT, Franco-Arellano B, Schermel A *et al.* (2017) Healthfulness and nutritional composition of Canadian prepackaged foods with and without sugar claims. *Appl Physiol Nutr Metab* **42**, 1217–1224.
49. Krukowski RA, Sparks C, DiCarlo M *et al.* (2013) There's more to food store choice than proximity: a questionnaire development study. *BMC Public Health* **13**, 586.
50. Nielsen (2016) *Understanding the Top Drivers Behind Shoppers' Store Choices*. https://www.nielsen.com/us/en/insights/article/2016/understanding-the-top-drivers-behind-shoppers-store-choices (accessed 2019 July 24).
51. Tugault-Lafleur CN & Black JL (2019) Differences in the quantity and types of foods and beverages consumed by Canadians between 2004 and 2015. *Nutrients* **11**, 526.
52. Chaix B (2018) Mobile sensing in environmental health and neighborhood research. *Annu Rev Public Health* **39**, 367–384.
53. Chen C, Ma J, Susilo Y *et al.* (2016) The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp Res Part C Emerg Technol* **68**, 285–299.
54. Suárez-Vega R, Gutiérrez-Acuña JL & Rodríguez-Díaz M (2015) Locating a supermarket using a locally calibrated Huff model. *Int J Geogr Inf Sci* **29**, 217–233.
55. Wang YC & DeSalvo K (2018) Timely, granular, and actionable: informatics in the public health 3.0 era. *Am J Public Health* **108**, 930–934.
56. Khoury MJ, Armstrong GL, Bunnell RE *et al.* (2020) The intersection of genomics and big data with public health: opportunities for precision public health. *PLOS Med* **17**, e1003373.

https://doi.org/10.1017/S1368980021003645 Published online by Cambridge University Press