SSM - Population Health 17 (2022) 101001

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Behavioral factors are perhaps more important than income in determining diet quality in Canada

Seyed H. Hosseini [a], Marwa Farag [b], Seyedeh Zeinab Hosseini [a], Hassan Vatanparast [a,b,*]

[a] College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK S7N 4Z2, Canada
[b] School of Public Health, University of Saskatchewan, Saskatoon, SK S7N 4Z2, Canada

ARTICLE INFO

Keywords:
Diet quality
Income deciles
Behavioral factors
Immigration
Machine learning

ABSTRACT

This study examines the importance of income in determining the diet quality of Canadian adults measured based on Nutrient Rich Food Index version 9.3. We used the latest available data on Canadians' consumption of foods and nutrients from the Canadian Community Health Survey-Nutrition 2015. The Canada' Food Guide classification was used for categorizing food groups based on types of food and their healthiness. Unsupervised and supervised machine learning models were employed in order to examine the links between income and the choice of foods. We first employed cluster analysis to identify the dietary patterns among individuals included in the sample and then we examined whether the intakes of various food groups across the identified clusters vary by income levels. Further, we evaluated the association between diet quality and income using Lasso Regression to determine the most important predictors of diet quality among adults in Canada. The results of both cluster analysis and regularized regression model suggested that behavioral factors and cultural backgrounds are more important determinants of diet quality among adults in Canada.

1. Introduction

Despite the lack of a global definition for diet quality, the term diet quality is used to indicate the extent to which dietary recommendation are followed by individuals to reach best health status possible (Alkerwi, 2014). Four main aspects are captured through scanning the diet quality of individuals including adequacy, moderation, variety and balance (Garriguet, 2009). Hence, diet quality has been investigated using nutrients and/or food groups by performing different summary scores or indices against the dietary recommendations or a-priori diets (Alkerwi, 2014; Afshin et al., 2019). Typically a high quality diet includes fruits, vegetables, beans and legumes, nuts and seeds, whole grains, milk, total polyunsaturated fatty acids, fish, plant omega-3s, and dietary fibre; and low consumptions of unprocessed red meats, processed meats, sugar-sweetened beverages, saturated fat, trans fat, dietary cholesterol, and sodium (Afshin et al., 2019; Imamura et al., 2015).

Diet quality has been known to be an important factor linked to obesity and morbidity and mortality related to non-communicable diseases such as hypertension, type 2 diabetes, cardiovascular diseases, osteoporosis, and cancer, worldwide (Echouffo-Tcheugui & Ahima, 2019; Darmon & Drewnowski, 2008; Hiza et al., 2013). Socioeconomic factors including race, ethnicity, income and educational attainment play important roles in choice of diet (Schoufour et al., 2018; Darmon & Drewnowski, 2008). Income status is one of the Socioeconomic Status (SES) that is known to affect diet quality.

There are mixed findings about the link between income and diet quality. There are studies, suggesting that empty-calorie foods such as non-enriched refined grains (that are enriched in Canada and contains some key nutrients of public health), fats, or added sugars are cheaper to obtain. Hence, individuals with lower income are likely to consume more of these less-nutritious foods and people with greater affluence are likely to have a higher consumption of nutrient-dense foods (Darmon & Drewnowski,2008; Kirkpatrick et al., 2010; Kirkpatrick & Tarasuk, 2003; Olstad et al., 2021a). However, there are studies showing income is not among the most important factors determining food choices. (Carlson et al., 2014; Gao et al., 2013; Garriguet, 2009; Hiza et al., 2013). The weak link between income and diet quality has been explained by the abundance of foods resulting from the development of advanced technologies in agriculture sector (Philipson & Posner, 1999; Lakdawalla & Philipson, 2009; Lakdawalla et al., 2005; Cutler et al., 2003). These technologies include a combination of agronomic practices, improved varieties (Fischer et al., 2012) and the efficient use of

* Corresponding author. School of Public Health, University of Saskatchewan, Saskatoon, SK S7N 4Z2, Canada.
E-mail addresses: h.hosseini@usask.ca (S.H. Hosseini), marwa.farag@usask.ca (M. Farag), seh65@mail.usask.ca (S.Z. Hosseini), vatan.h@usask.ca (H. Vatanparast).

https://doi.org/10.1016/j.ssmph.2021.101001
Received 18 April 2021; Received in revised form 25 November 2021; Accepted 11 December 2021
Available online 17 December 2021

water resources and nitrogen (Sadras & Lawson, 2013) that have led to considerable yield gains such as Green Revolution (Evenson & Gollin, 2003).

The majority of Canadians follow a poor diet quality (Nshimyumukiza et al., 2018), and this is especially the case for younger individuals (Garriguet, 2009). In Canada, poor diet quality was ranked as the fifth-highest contributor to the chronic diseases burden and deaths in 2019 (Institute for Health Metrics and Evaluation, 2019). The new Healthy Eating Strategy proposed by Health Canada aims to improve the diet quality of Canadians through more informed choices in choosing foods, improving foods' nutrition qualities, protecting those people who are vulnerable, and improving access to nutritious foods (Health Canada, 2016). Therefore, knowing about the factors affecting food choices could contribute to setting more effective policies.

Considering the mixed results in the previous literature regarding the link between income and diet quality, the objective of this study is to examine the links between income and the diet quality and how it compares to other SES factors among Canadian adults (19-70y) using the latest available nationally representative data.

To address the objective of this study, we first employ cluster analysis to identify the dietary patterns among individuals included in the sample and then examine whether the intakes of various food groups across the identified clusters vary by income levels. Further, this study investigates the association between diet quality and income using Lasso Regression (LR) method.

2. Data and method

2.1. Data

2.1.1. The Canadian Community health survey

The CCHS 2015-Nutrition is a cross-sectional survey representing Canadians both nationally and provincially. This survey provides a comprehensive assessment of Canadians' dietary intake and includes information about participants' socio-demographics and socioeconomic characteristics. The CCHS sample size is 20,487 and the sampling frame of CCHS included "clusters" that were created from Census Dissemination Areas (DAs). Each DA generally contains 400 to 700 people. To achieve a sample representing Canada's population in the cases of socioeconomic status, sex, age and geography, a three-stage sampling design was employed (Statistics Canada, 2017). The CCHS 2015 covers 98% of the of Canada's population and has the response rate of 62%. However, the survey does not represent the residents of Indigenous reserves, the residents of the three Canadian territories, and those living in institutions (correctional facilities, military bases, and hospitals). For this study, we used the public use microdata file of the CCHS-Nutrition, 2015; and for descriptive analyses, we used a sampling weight provided by Statistics Canada. (Statistics Canada, 2017). It should be also noted that, although, the CCHS administrated yearly, there are only two years during which dietary intakes data were collected that the first one was in 2004 and the second one was in 2015 and we used the most recently available data in 2015.

Following the recommendations made by Templ et al. (2020), the adjusted Box-Plot method was used to identify the outliers (Templ et al., 2020). The outliers were identified and excluded from our sample for energy intakes (<396 Kcal and >4889 Kcal), grams of food consumed (<752 g and >7142 g), NRF (<−92 and >1094) and BMI (<17.35 kg/m^2 and >47.5 kg/m^2). Pregnant and breast-feeding women were excluded from our sample because it is likely that their diets are affected by their status. Therefore, our sample includes 10,196 individuals representing 22,594,044 adults aged 19–70 years old in Canada.

The age and sex groups chosen in this study are based on Dietary Reference Intake (DRI) age and sex groups. The best available source for Canadian benchmark is Statistics Canada' 2016 census profile of Canada (Statistics Canada, 2017). However, the age and sex groups reported by Statistics Canada are slightly different than upper and lower bounds of DRI reference age and sex groups. Therefore, to indicate if CCHS properly represents Canada population, Table 1, shows sample size, weighted values, and Canadian Benchmark for various SES.

2.1.2. Dietary intake data

In CCHS Nutrition-2015, the dietary intake data were collected via two 24-h recall method. The first 24-h recall was implemented for all participants to achieve a better estimation of distributions of usual intakes a third of respondents were randomly selected for the second 24-h recall. We, however; used the first recall as the mean intakes from one recall is similar for statically adjusted mean intakes from two recalls. The unit of food intake in our study is the number of servings of foods consumed. In CCHS, Nutrition 2015 food intakes have been categorized based on different classifications, among which we chose the one used in 2007 Canada's Food Guide (CFG) (Health Canada, 2018). In the CFG's classification, there are four primary food groups, including grain products, milk and alternatives, meat and alternatives, and finally vegetables and fruit (Health Canada, 2014). In addition, under each primary category, there are sub-categories. For instance, the grain products include whole grains; refined grains, enriched; and refined grains, not enriched categories.

Furthermore, a series of thresholds for fat, sugars, sodium, and saturated fats were used to classify the foods to distinguish between food groups based on their healthiness (Health Canada, 2014). The lower thresholds for fat, sugars, and sodium were 3 g per reference amount, 6 g per reference amount, and 140 mg per reference amount, respectively. The upper threshold for fat, sugars, sodium, and saturated fat were 10 g per reference amount, 19 g per reference amount, 360 mg per reference amount, and 2 g per reference amount, respectively. The foods that did not exceed any of the lower cutoffs were categorized as Tier 1 foods. The foods that exceeded one or two of the lower thresholds and did not exceed the upper cutoffs were considered as Tier 2 foods. Tier 3 foods included those foods whose fat, sugars, and sodium surpassed the lower cutoffs without exceeding the upper levels or only one of the upper thresholds. Finally, the foods that exceeded at least two upper cutoffs were considered as Tier 4 foods.

The number of food groups initially was 74 groups, and after removing those food groups that less than 1 percent of adults in our

Table 1
Comparison of CCHS sample, weighted data to Canadians national benchmarks.

Variable	CCHS sample	Weighted data	Benchmark	Source of Benchmark
Male (Age: 18 years and over)	6633	13,791,001	14,129,851	Statistics Canada (2016 census profile of Canada)
Female (Age: 18 years and over)	7642	14,540,730	14,566,754	Statistics Canada (2016 census profile of Canada)
Adults with Secondary (high) school diploma or equivalency certificate (Age: 15 years and over)	3966	7,533,861	7,576,400	Statistics Canada (2016 census profile of Canada)
Bachelor's degree or university certificate/ diploma/degree above Bach level (Age: 15 years and over)	3381	7,835,635	6,659,620	Statistics Canada (2016 census profile of Canada)
Immigrants (Age: 15 years and over)	3072	7,891,193	7,540,830	Statistics Canada (2016 census profile of Canada)

sample consumed, we left with 55 food groups. It should also be noted that the daily numbers of servings consumed by each respondent in our sample were adjusted for 2000 Kcal of daily energy intakes.

2.1.3. Diet quality index

In this study, we have also calculated and used the Nutrient Rich Food Index (NRF) version 9.3 (Fulgoni et al., 2009) for Canada. The NRF 9.3 represents a measure of diet quality based on the recommended daily consumption of 12 nutrients. The NRF 9.3 allocates positive weights to the consumption of nine nutrients, including protein, fiber, vitamin C, vitamin A, vitamin D, iron, calcium, magnesium, and potassium. The intakes of three components, including sodium, sugars, and saturated fatty acids, are associated with a series of negative weights in the calculation of the NRF 9.3. Following Barr et al. (2018), we made adjustments to generate the NRF 9.3 index for Canadian data (Barr et al., 2018). For instance, we adjusted nutrient intakes for 2000 Kcal of energy intake and replaced vitamin E with vitamin D.

A series of nutrients selected in NRF 9.3 are based on FDA's definition of "healthy foods" including foods with fibre, protein, vitamin C, vitamin A, iron and calcium (Fulgoni et al., 2009). The additional nutrients including vitamin D, potassium, and magnesium are included in the calculations because they are identified as the nutrients of concern by Health Canada (Health Canada, 2014).

As it is indicated in the definition of NRF (Drewnowski & Fulgoni III, 2008, page 32), NRF is a unitless index. For each person, a score is allocated to each of the 12 nutrients. The score's nominator is the amount of a nutrient consumed by an individual and its denominator is the daily recommended value of that nutrient weighted by the individual energy intakes in relation to 2000 Kcal of energy. For instance, in the case of protein the daily recommended value of protein is 50 g per day, therefore if an individual consumes 70 g of protein and his/her energy intakes is 1800 Kcal, the individual's protein score will be equal to 100*(70)/(50*(1800/2000)) = 150 that is a unitless score. Final calculation of NRF 9.3 includes adding the scores of nine encouraged nutrients and subtracting the scores of three discouraged nutrients.

2.1.4. Income measure and other variables

The primary measures of income in our analyses are income deciles. The income deciles reported in our study are calculated by Statistics Canada and account for the province of residence and household size. There were missing values in the incomes reported by participants of CCHS, therefore, Statistics Canada imputed the total household income for about a fifth of adults using nearest neighbor donor approach (Statistics Canada, 2017). The imputation is based on the structure of family, type of residence, socioeconomic status, the median of postal code tax, household size and health variables (Yeung & Thomas, 2013).

Other variables included in our analyses are as follows: age-sex groups (males or females aged 19–30, 31–50, and 51–70 years old), the highest level of education achieved by a member of the household (university degree and less than university degree), smoking (smoker or non-smoker), physical activity (whether the respondent had more than 150 min of physical activity per week or not), marital status (single, never married, widowed or separated and married or common-law partner) and immigration status (immigrant or non-immigrant). The aforementioned variables, along with income deciles and 55 food groups, were used as the predictor variables of regularized regression model.

2.2. Method

2.2.1. Dietary pattern assessment using cluster analysis

K-mean cluster analysis was used to identify dietary patterns across income groups (Jain, 2010; Newby & Tucker, 2004). K-mean cluster analysis is a reiterative process classifying data into K groups. The procedure starts with the choice of K random values in the data. The observations are first allocated to the K randomly chosen values based on their distances to these random values. Subsequently, assigning all data points to the K initial centroids, an average for each cluster is calculated and again, all data points' distances to the mean value of previous clusters are calculated to form new clusters. This process continues until no further changes can be identified in terms of proximity to the mean values of the latest clusters. To determine the optimal number of clusters, we used Calinski and Harabasz index (Caliński & Harabasz,1974) indicated to be the most effective method among 30 stopping rules in identifying the correct number of clusters (Milligan & Cooper, 1985).

2.2.2. Regularized regression: Lasso regression

Examining the factors predicting the diet quality of Canadian adults, we employed LR. The LR belongs to the family of linear regularized regression models. The Least Square regression (LSR) minimized the Residuals Sum of Squares (RSS). The RSS with p predictors can be written as $RSS = \sum_{i}^{n}(y_i - b_0 - \sum_{j=1}^{p} b_j x_{ij})$. However, the LR model has an extra term as follows $RSS + \lambda \sum_{j=1}^{p} \left| b_j \right|$. The term $\lambda \sum_{j=1}^{p} \left| b_j \right|$ is known as shrinkage penalty in which the λ is the tunning parameter that can be determined in the statistical learning process. The choice of λ can be based on various model metrics such as minimum Root Mean Square Error (RMSE). If $\lambda = 0$, the LR model will be equivalent to LSR and as λ increases, the coefficients shrink toward zero. The LR has some advantages over LSR model. The LR estimates, have lower variance at the cost a small amount of bias compared to LSR estimates. In addition, because of the presence of penalty in the model, the LR keeps the most important variables and shrink the coefficients of less important variables to zero. Furthermore, as the penalty term in the LR model controls for multicollinearity between the predictors, we will be able to include all levels of nominal predictors in the form of dummy variables and allow the model to choose the important predictors.

The data was split into training and testing sets to evaluate the predictability of our model. Finding the optimal λ minimizing RMSE, we used 10-fold cross-validation.

3. Results

Fig. 1 indicates the weighted prevalence of a set of behavioral and socioeconomic factors of Canadian adults aged 19–70 years old across 10 income deciles. Income-based differences, especially between low- and high-income deciles, exist in marital status, immigration status, higher education, smoking habit, and physical activity. In addition, the prevalence of individual with diabetes is considerably higher across low-income deciles compared to high-income deciles. Furthermore, the proportion of females aged 19 to 30 are higher across low-income deciles.

3.1. Cluster analysis results

The optimal number of clusters for 55 food groups consumed by the sample respondents is two clusters. Conducting cluster analysis, we observed that the average NRF 9.3 across identified clusters are 478.1 ± 3.8 and 245.5 ± 2.0. Therefore, the cluster including individuals with an average NRF 9.3 of 478.1 is called the High Quality (HQ) cluster, and the other one is called the Low Quality (LQ) cluster.

Fig. 2 indicates the weighted frequencies of various behavioral and socioeconomic factors across HQ and LQ clusters. As we can see, the prevalence of males and especially younger males is higher in LQ cluster in comparison with HQ cluster. In addition, we can see two important patterns where 41.6% of individuals in HQ cluster are immigrants and 12% of individuals in HQ clusters are smokers.

Table 2 shows the average intakes of nutrient used in the calculation of NRF 9.3 along with average energy and grams of foods intakes of the

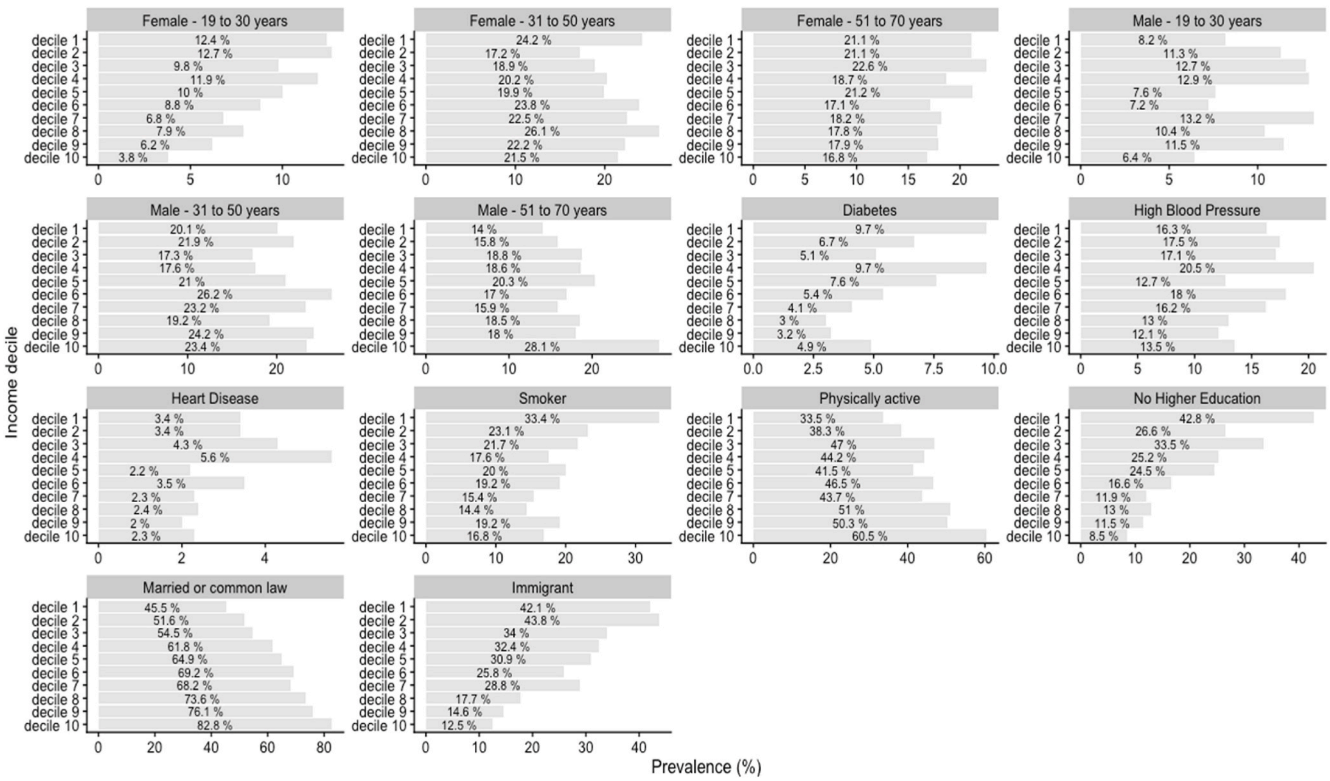

Fig. 1. The weighted frequencies of a various behavioral and socioeconomic factors across income deciles for Canadian adults 19–70 years old.

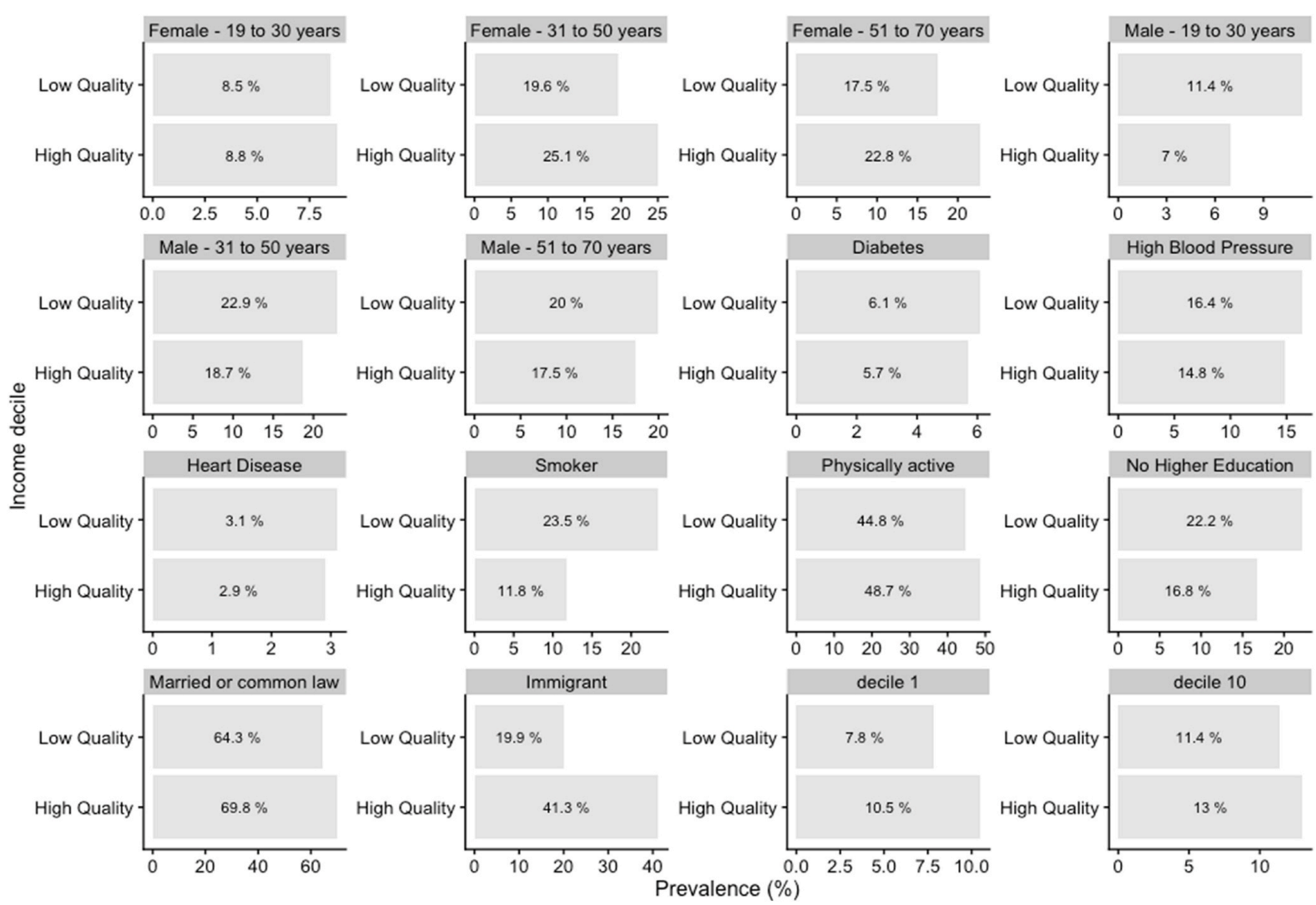

Fig. 2. The weighted frequencies of various behavioral and socioeconomic factors across HQ and LQ clusters.

Table 2
The average intakes of nutrient used in the calculation of NRF along with average energy and grams of foods intakes of the participants in the two clusters.

Nutrients	High Quality Cluster	Low Quality Cluster	P-value
Calcium intake from food sources (mg)	729.3	830.4	<0.001
Potassium intake from food sources (mg)	2823.2	2638.7	<0.001
Magnesium intake from food sources (mg)	338.3	290.6	<0.001
Iron intake from food sources (mg)	11.4	12.9	<0.001
Protein intake from food sources (g)	78.7	81.5	<0.001
Total dietary fibre intake from food sources (g)	20.3	15.3	<0.001
Vitamin A from food in retinol activity equiv. (mcg)	707.2	594.4	<0.001
Vitamin D intake from food sources (mcg)	5.1	4.7	<0.001
Vitamin C intake from food sources (mg).	110.9	84	<0.001
Sodium intake from food sources (mg)	2275.9	3044	<0.001
Total saturated fatty acid intake from food (g)	17.7	26.1	<0.001
Total sugars intake from food sources (g)	79.4	92.2	<0.001
Energy intake from food sources (Kcal)	1676.7	2017.1	<0.001
Amount of food (g)	3000.9	2956.3	<0.001

g: gram; mg: milligram; mcg: microgram, Kcal: kilo calories.

participants in the two clusters (HQ and LQ). Except for calcium, the participants of HQ clusters have significantly higher intakes of nutrients whose intakes are encouraged. However, the participants of LQ clusters have significantly higher intakes of sodium, saturated fatty acid, and sugars than the participants of HQ clusters.

Table 3 illustrates the average number of servings of various food groups consumed by the participants in HQ and LQ clusters. The participants of HQ clusters generally have higher intakes of healthier foods (Tier 1 and Tier 2 food groups) except in the case of grain, non-whole enriched foods (that are fortified by iron, folate, thiamin, niacin, and other nutrients in Canada), and fruit juice that can be classified as relatively less healthier foods. However, the participants of LQ clusters on average consume more of Tier 3 and Tier 4 food groups.

Fig. 3 shows the average number of servings of the five top foods consumed across HQ and LQ clusters in the case of grains, fruits and vegetables, meat and alternatives and milk and alternatives food groups. In general, we can observe that the being in HQ cluster contains higher consumption of whole grains, fruits, and vegetables, poultry, and fish in comparison with LQ pattern. However, being in LQ cluster is associated with higher consumption of non-whole grains, beef, and processed meat products. Furthermore, the intakes of foods that are classified as Tier 3 and Tier 4 are more frequent in LQ pattern.

Examining the average number of servings of foods consumed across income deciles in LQ and HQ clusters in Fig. 4 we can see that, in most of the cases there is no statistically or clinically significant differences in the consumption of food groups across income deciles. In other words, those individuals following LQ or HQ dietary patterns do not change their eating behaviors as their income class changes.

However, there are few cases among which the intakes of food groups are associated with income. For instance, the consumption of whole grain Tier 2 decreases by more than 0.75 servings from the bottom income decile to top income decile. Similarly, the intake of dark green vegetables Tier 1 is associated with the income levels for more or less both clusters, although, the changes are not significantly higher between the first and the last deciles of income in terms of the number of servings consumed. Furthermore, we can observe that the intakes of

Table 3
The average number of servings of various food groups consumed by the participants in High-Quality and Low-Quality clusters.

Food Group	High Quality Cluster (Servings/day)	Low Quality Cluster Servings/day	P-value
Fruit, No Juice, Tier 1	2.3	0.9	<0.001
Fruit, No Juice, Tier 3	0.1	0	<0.001
Fruit, Juice, Tier 2	0.4	0.5	0.02
Vegetable, Dark Green Tier 1	1.1	0.3	<0.001
Vegetable, Orange, Tier 1	0.4	0.1	<0.001
Vegetable, Potato, Tier 1	0.4	0.4	0.3
Vegetable, Potato, Tier 3	0.1	0.4	<0.001
Vegetable, Other, Tier 1	2.0	0.9	<0.001
Vegetable, Other, Tier 3	0.1	0.2	<0.001
Grain, Whole, Tier 1	0.5	0.1	<0.001
Grain, Whole, Tier 2	1.4	0.5	<0.001
Grain, Whole, Tier 3	0.1	0	<0.001
Grain, Whole, Tier 4	0.1	0.1	0.3
Grain, Non-Whole, Enriched Tier 1	0.4	1.1	<0.001
Grain, Non-Whole Enriched Tier 2	1.2	2.7	<0.001
Grain, Non-Whole, Enriched Tier 3	0.4	0.7	<0.001
Grain, Non-Whole, Enriched Tier 4	0.2	0.4	<0.001
Meat, Beef, Tier 3	0.3	0.5	<0.001
Meat, Poultry, Tier 2	0.6	0.2	<0.001
Meat, Poultry, Tier 3	0.3	0.2	<0.001
Meat, Processed, Tier 3	0.1	0.2	<0.001

Tier 1 refers to the healthiest foods in terms of having lower amount of sodium, sugar and saturated fatty acid and Tier 4 refers to the foods that have significantly high amounts of sodium, sugar, and saturated fatty.

fluid milk Tier 1 increases and the consumptions of milk fluid Tier 2 decrease as income increases for both clusters.

3.2. Lasso regression results

The optimal λ, minimizing RMSE was significantly close to zero (i.e., 1e-10), implying that the LR estimates are close to LSR estimates. We evaluated the performance of LR using the training set and applied the results to the testing set, including 80% and 20% of the sample, respectively. Fig. 5 illustrates the scatterplot of predicted NRF 9.3 versus the actual values of NRF 9.3 in the testing set. As we can see, the LR has a desirable performance in predicting the values of NRF 9.3 such that the coefficient of the fitted line in Fig. 5 is 0.67.

Fig. 6 indicates the coefficients of LR model for food groups and behavioral and socioeconomic predictors. As we can see, the top five food groups with a positive effect on NRF 9.3 are fruits and vegetables mostly categorized as Tier 1, including fruit Tier 1, dark green vegetables Tier 1, orange vegetable Tier 1, fruit juice Tier 2 and other vegetables Tier 1. The food groups with a negative effect on NRF 9.3 belong to Tier 4 and Tier3 categories and include non-whole grain, enriched Tier 4, processed meat Tier 4, milk-based foods other than fluid milk Tier 4 and non-whole grain, not-enriched Tier 4. Among the behavioral and socioeconomic factors, being female aged 51–70 years old, being in top three income deciles, having someone with higher education in the household, being non-smoker, and being immigrant have a positive association with diet quality. However, being less physically active, being married or being in common law and being in the bottom income deciles have a negative association with the diet quality of adults in Canada.

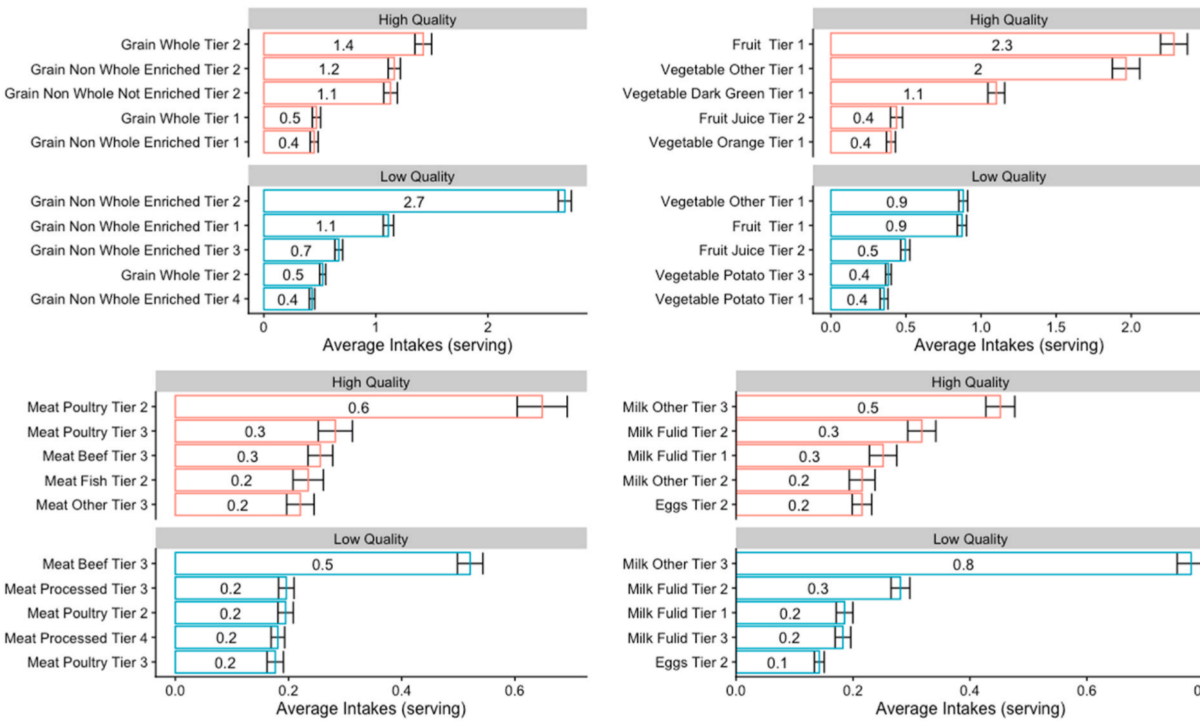

Fig. 3. Average number of servings per day of top 5 foods consumed by Canadian adults aged 19–70 years old across clusters.

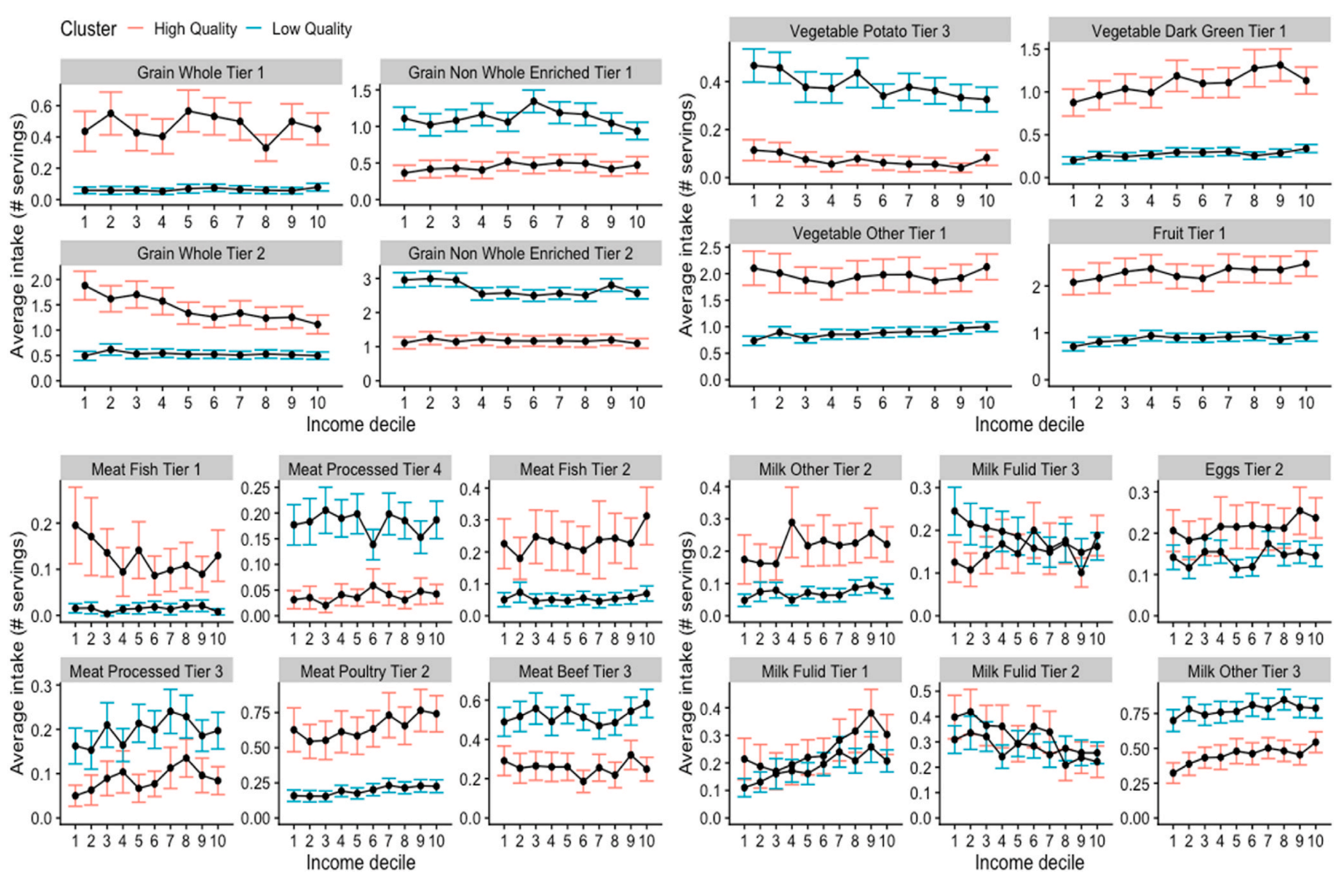

Fig. 4. Average numbers of servings per day of various foods consumed by Canadian adults aged 19–70 years old across income deciles in High-Quality and Low-Quality clusters.

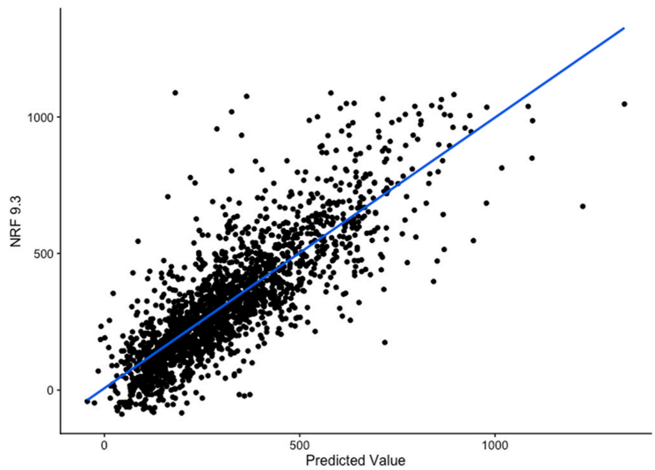

Fig. 5. The Lasso regression performance (predicted NRF 9.3 versus the real values of NRF 9.3 in the testing set).

4. Discussion

In this study, we evaluated the links between income, dietary intakes, and diet quality among Canadian adults. The CFG's Tiers classification was used to identify dietary patterns of adults in the sample. The cluster analyses led to the identification of two clusters called HQ and LQ. The HQ dietary pattern was associated with higher intakes of fruits and vegetables, whole grain, and poultry, compared to higher consumption of non-whole grains, beef, and less intakes of fruits and vegetables in LQ cluster. Furthermore, the individuals in HQ pattern consumed more of Tier 1 and Tier 2 foods, while those categorized in LQ pattern had considerable intakes of Tier 3 and Tier 4 foods. The prevalence of immigrants, non-smokers, and females aged 31–50 and females aged 51–70 were considerably higher in HQ dietary patterns. Furthermore, except few cases, such as whole grain Tier 2 and dark green vegetables Tier 1, no statistical or clinical differences were observed in the average number of servings of foods consumed across income deciles.

Examining the factors determining the diet quality of Canadian adults, an LR model was estimated. Machine learning was used to tune the model, and the results implied that the LR model is very close to LSR model. The consumption of fruits and vegetables in Tier 1 and Tier 2 positively impacted diet quality. Among the behaviors and socioeconomic factors, the behavioral factors such as being physically less active and being non-smoker have a higher importance in predicting deity quality measures compared to sex-age group, household income, and household education.

As it was discussed earlier, in general, there are two primary theories about the relationship between income and food choices. The first group of studies implies that individuals with lower income levels consume more empty-calories foods because they are cheaper (Darmon & Drewnowski,2008; Drewnowski & Specter, 2004). The second set of studies suggests that due to advancements in agricultural technologies food prices have decreased over time and specially in technologically advanced countries food is available for the majority of people (Philipson & Posner, 1999; Lakdawalla & Philipson, 2009; Lakdawalla et al., 2005) implying that the link between income and diet quality is not significant. Our findings are consistent more with the second strand of literature such that the importance of income is not as much as the importance of behavioral factors.

The existing empirical works support both theories. An extensive review of the empirical results indicates individuals with greater affluence are likely to have a higher consumption of nutrient-dense foods (such as lean meat, whole grain, and seafood) (Darmon & Drewnowski, 2008). However, there are studies showing there is no connection between income and food choices. Gao et al. (2013) showed that an increase in income does not affect the demand for diet quality in the U.S (Gao et al., 2013). Carlson et al. (2014), indicated that while income has

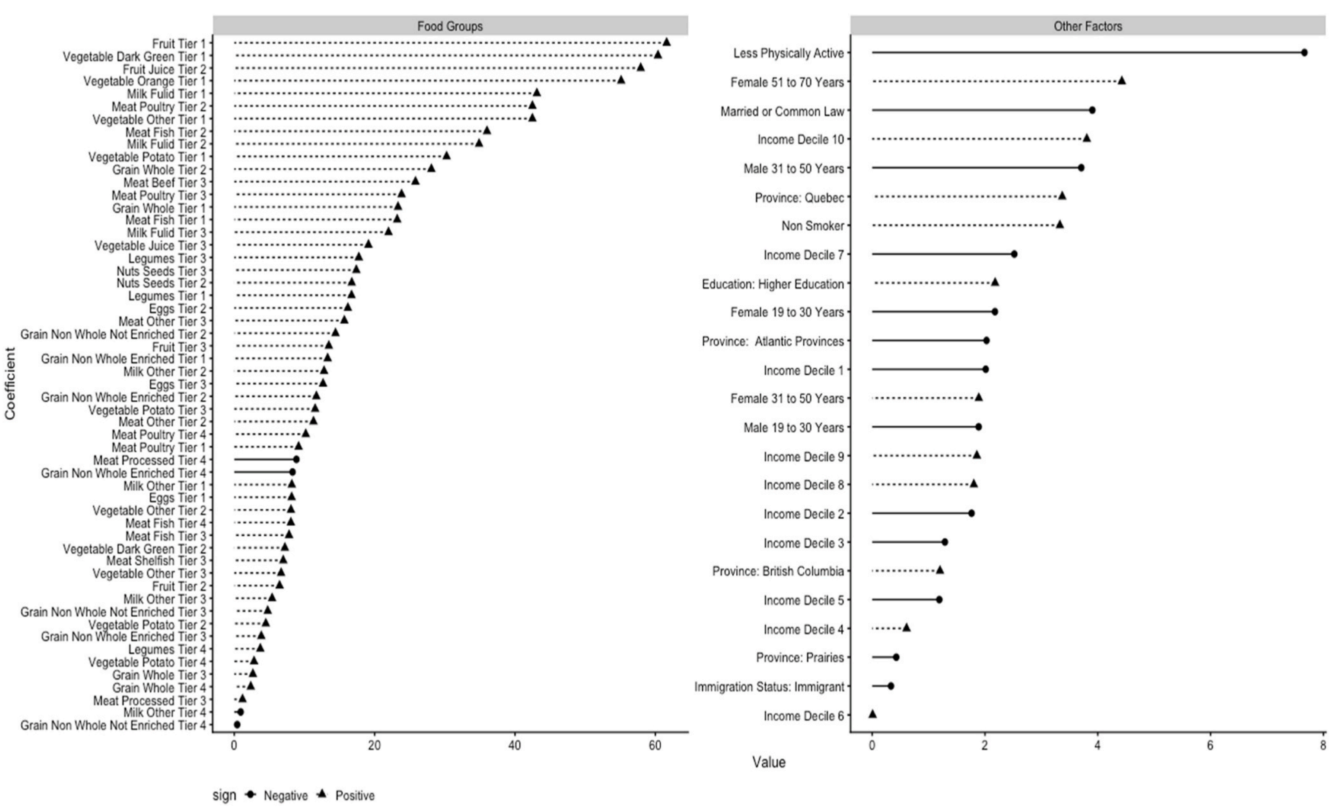

Fig. 6. The coefficient plots of Lasso regression results for food groups, behavioral and socioeconomic predictors.

small and positive impact on diet quality, the effect of other socio-economic factors including education, immigration status, and sex are considerably higher on the diet quality of Americans (Carlson et al., 2014). Moreover, Hiza et al. (2013), showed that in comparison with income, cultural background, ethnicity and sex are more important factors in determining diet quality in the U.S (Hiza et al., 2013).

In the case of Canada, Olstad et al. (2021) showed socioeconomic positions including education and income significantly explain the differences in diet quality of adults (Olstad et al., 2021a) and children (Olstad et al., 2021b). Kirkpatrick and Tarasuk (2003) indicated that in general low-income households spent less on the purchase of fruits and vegetables (Kirkpatrick & Tarasuk, 2003). Ricciuto and Tarasuk (2007) found significant differences in the nutritional quality of foods purchased across income levels (Ricciuto & Tarasuk,2007). However, Garriguet (2009), reported small differences in a measure of diet quality across income levels (Garriguet, 2009).

Understanding the choice of diet quality and income, one can take into account the studies examining the effectiveness of the Supplemental Nutrition Assistance Program (SNAP). The SNAP can be considered as a direct cash transfer (Anderson & Butcher, 2016; Alston et al., 2009). It has been indicated that the SNAP increased the food expenditures of the program's participants by 20% (Unnevehr et al., 2010). However, little evidence was found on the effect of such a program on the diet quality of participants (Fox et al., 2004). In other words, it seems that SNAP was not successful to improve the diet quality of participants, even if it provided a financial means for provision of nutrient-dense items. It is discussed that behavioral factors such as self-control problem, the default option, mental accounting and perception of fixed cost versus variable costs can explain why food assistance programs in the U.S were not successful in improving the diet quality of participants (Just et al., 2007).

The higher prevalence of immigrants, non-smokers, and females ages 31–50 and 51–70 were in HQ dietary patterns stem from the fact that dietary habits are mainly shaped by factors such as culture, environment, and education. The immigrants are usually healthier and have better diet quality than Canadians. This phenomena known as the "Healthy Immigrant Effect" stems from screening procedures of Canada's government limiting the admission of less healthy applicants and the fact that healthier people are more likely to have the means of immigration (Sanou et al., 2014).

The positive association between smoking and poor diet quality could be related to a smoker's brain reactions to smoking that affect his/her mood and appetite (Fowler et al., 1996). In addition, poorer diet quality of smokers in comparison with non-smokers could be related to fact that, some people in general have unhealthy lifestyles (Dallongeville et al., 1998; MacLean et al., 2018; Alkerwi et al., 2017).

The higher quality of diets of females ages 31–50 and 51–70 could be linked with gender differences in food choices. That is, women do care more about healthy nutrition, and are more concerned and are more under pressure about controlling their body weights stemming from social pressure (Grzymisławska et al., 2020).

In addition, the effect of behavioral factors on the choice of a healthy lifestyle could be explained by the findings of several studies indicating low earning causes stress and lowers productivity. Therefore, the poor seem to have a higher rate of time preference, implying they prefer the present more because of all the risks and uncertainty they face. These two characteristics in turn, force the poor to limit their attention to the present and value their current habit at the cost of their goals and future (Haushofer & Fehr, 2014). In this case, higher levels of education, sex or cultural background could offset the higher values that less affluent individuals assign to the present.

Therefore, policies aiming at improving the diet quality of individuals in Canada may not be merely focused on providing financial means for individuals, so they adopt healthier diet.

Considering the results of our analyses, designing effective policies in improving diet quality requires collective actions (Olstad, Campbell and Raine 2019) that expands access to post-secondary education and other means lowering the rate of time preference for those in lower income deciles. Also it should be noted that as the individuals are affected by the environment they live in, a policy should be designed in a way to addresses the connection between individual's choices and environments around them (Olstad et al., 2021a).

There are some limitations in our study. First, income and dietary intake data, are self-reported. The self-reported household income used in this study may be an imperfect measure of perceived economic means to purchase adequate food for the household. Since other economic measures may play a role in the context of food choice, such as financial stress or difficulty in rental payment. We also did not have access to the data on food expenditures. Hence, we could not have a more thorough analysis of the demand for healthy versus unhealthy foods across income levels. However, we employed different statistical methods to provide more accurate analyses of the consumption patterns and their association with income. Finally, One of the limitations of CCHS 2015-Nutrition data is that they do not provide information about a wide range of factors affecting the diet quality such as access to public transportation, the placement of supermarkets in neighborhoods (i.e., are there supermarkets in lower-income neighborhoods to the same extent as higher-income neighborhoods. (Ver Ploeg and Rahkovsky 2016). Therefore, our analyses are limited to the existing information.

It should also be noted that this study benefits from large sample size and the use of different statistical methods to meet its objective in providing a more robust answers to the research questions. In addition, this study examined the relationship between income, diet, and health outcomes from a multidisciplinary perspective among Canadians.

5. Conclusion

Using the most recent nationally representative data, we applied both supervised and unsupervised machine learning models to examine the links between income and the diet quality and how it compares to other SES factors among Canadian adults (19-70y). Canadians with high quality diet consumed healthier foods independent to their income status. Immigrants, non-smokers, and women aged 31 and over had considerably greater presence in high quality dietary patterns group. Overall, in comparison with income, we found behavioral factors and cultural dietary practices are more important determinants of diet quality among adults in Canada. Choosing a healthier diet by individuals is made in a multifaceted process that can be taken into consideration in planning policies and health promotion initiatives aiming at improving diet quality in Canada.

Author statement

Seyed Hamzeh Hosseini: Conceptualization, Data curation, Formal analysis, Software, Writing- Original draft, Visualization, Methodology, Validation,; Writing - review & editing.

Marwa Farag: Conceptualization, Writing- Original draft, Writing - review & editing, validation.

Seyedeh Zeinab Hosseini: Conceptualization, Writing- Original draft, Writing - review & editing, validation.

Hassanali Vatanparast: Conceptualization, Data curation, Formal analysis, Project administration, Supervision, Validation, Writing - original draft; Writing - review & editing.

Source of support

Mitacs Canada, Alberta and Saskatchewan Wheat Development Commissions, Grain Farmers of Ontario grant numbers 420066-8 and "The APC was funded by Cereals Canada Inc."

Declaration of competing interest

The authors declare no conflict of interest.

References

Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., Salama, J. S., Mullany, E. C., Abate, K. H., Abbafati, C., Abebe, Z., Afarideh, M., Aggarwal, A., Agrawal, S., Akinyemiju, T., Alahdab, F., Bacha, U., Bachman, V. F., Badali, H., Badawi, A., … Murray, C. J. L. (2019). Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the global burden of disease study 2017. *The Lancet, 393*(10184), 1958–1972.

Alkerwi, A. (2014). Diet quality concept. *Nutrition, 30*(6), 613–618.

Alkerwi, A., Baydarlioglu, B., Sauvageot, N., Stranges, S., Lemmens, P., Shivappa, N., & Hébert, J. R. (2017). Smoking status is inversely associated with overall diet quality: Findings from the ORISCAV-LUX study. *Clinical Nutrition, 36*(5), 1275–1282.

Alston, J. M., Mullally, C. C., Sumner, D. A., Townsend, M., & Vosti, S. A. (2009). Likely effects on obesity from proposed changes to the US food stamp program. *Food Policy, 34*(2), 176–184.

Anderson, P. M., & Butcher, K. F. (2016). *The relationships among SNAP benefits, grocery spending, diet quality, and the adequacy of low-income families*. Washington, DC: Center on Budget and Policy Priorities.

Barr, S., Vatanparast, H., & Smith, J. (2018). Breakfast in Canada: Prevalence of consumption, contribution to nutrient and food group intakes, and variability across Tertiles of daily diet quality. A study from the international breakfast research initiative. *Nutrients, 10*(8), 985.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods, 3*(1), 1–27.

Carlson, A., Dong, D., & Lino, M. (2014). Association between total diet cost and diet quality is limited. *Journal of Agricultural and Resource Economics, 39*(1), 47–68.

Cutler, D. M., Glaeser, E. L., & Shapiro, J. M. (2003). Why have Americans become more obese? *The Journal of Economic Perspectives, 17*(3), 93–118.

Dallongeville, J., Marécaux, N., Fruchart, J.-C., & Amouyel, P. (1998). Cigarette smoking is associated with unhealthy patterns of nutrient intake: A meta-analysis. *Journal of Nutrition, 128*(9), 1450–1457.

Darmon, N., & Drewnowski, A. (2008). Does social class predict diet quality? *American Journal of Clinical Nutrition, 87*(5), 1107–1117.

Drewnowski, A., & Fulgoni, V., III. (2008). Nutrient profiling of foods: Creating a nutrient-rich food index. *Nutrition Reviews, 66*(1), 23–39.

Drewnowski, A., & Specter, S. E. (2004). *Poverty and obesity: The role of energy density and energy costs*, (1). Available at: https://academic.oup.com/ajcn/article-abstract/79/1/6/4690070.

Echouffo-Tcheugui, J. B., & Ahima, R. S. (2019). Does diet quality or nutrient quantity contribute more to health? *Journal of Clinical Investigation, 129*(10), 3969–3970.

Evenson, R. E., & Gollin, D. (2003). Assessing the impact of the green revolution, 1960 to 2000. *Science, 300*(5620), 758–762.

Fischer, T., Byerlee, D., & Edmeades, G. (2012). Crop yields and food security: Will yield increases continue to feed the world. In *Proceedings of the 12th Australian agronomy conference* (pp. 14–18).

Fowler, J. S., Volkow, N. D., Wang, G.-J., Pappas, N., Logan, J., Shea, C., Alexoff, D., MacGregor, R. R., Schlyer, D. J., & Zezulkova, I. (1996). Brain monoamine oxidase A inhibition in cigarette smokers. *Proceedings of the National Academy of Sciences, 93*(24), 14065–14069.

Fox, M. K., Hamilton, W. L., & Lin, B.-H. (2004). *Effects of food assistance and nutrition programs on nutrition and health*.

Fulgoni, V. L., Keast, D. R., & Drewnowski, A. (2009). Development and validation of the nutrient-rich foods index: A tool to measure nutritional quality of foods. *Journal of Nutrition, 139*(8), 1549–1554.

Gao, Z., Yu, X., & Lee, J. Y. (2013). Consumer demand for diet quality: Evidence from the healthy eating index. *The Australian Journal of Agricultural and Resource Economics, 57*(3), 301–319.

Garriguet, D. (2009). Diet quality in Canada. *Health Reports, 20*(3), 41.

Grzymisławska, M., Puch, E. A., Zawada, A., & Grzymisławski, M. (2020). Do nutritional behaviors depend on biological sex and cultural gender? *Advances in Clinical and Experimental Medicine: Official Organ Wroclaw Medical University, 29*(1), 165–172.

Haushofer, J., & Fehr, E. (2014). On the psychology of poverty. *Science, 344*(6186), 862–867.

Health Canada. (2014). Health Canada's proposed changes to the core nutrients declared in the Canadian nutrition facts table. Available at: https://www.canada.ca/en/health-canada/services/food-nutrition/public-involvement-partnerships/technical-consultation-proposed-changes-core-nutrients-declared-canadian-nutrition-facts-table/consultation.html. (Accessed 12 November 2021).

Health Canada, H. (2014). *The development and use of a surveillance tool: The classification of foods in the Canadian nutrient file according to eating well with Canada's food Guide*. Ontario): Health Canada Ottawa.

Health Canada's healthy eating strategy. (2016). Ottawa: Government of Canada. Available: https://www.canada.ca/en/services/health/campaigns/vision-healthy-canada/healthy-eating.html.

Health Canada. (2018). 2015 Canadian community health survey - nutrition - food and nutrition surveillance - health Canada. *Aem*. Available at: https://www.canada.ca/en/health-canada/services/food-nutrition/food-nutrition-surveillance/health-nutrition-surveys/canadian-community-health-survey-cchs/2015-canadian-community-health-survey-nutrition-food-nutrition-surveillance.html. (Accessed 4 August 2020).

Hiza, H. A. B., Casavale, K. O., Guenther, P. M., & Davis, C. A. (2013). Diet quality of Americans differs by age, sex, race/ethnicity, income, and education level. *Journal of the Academy of Nutrition and Dietetics, 113*(2), 297–306.

Institute for Health Metrics and Evaluation. (2019). *Country profile: Canada. Top 10 risks contributing to DALYs in 2019 and percent change, 2009–2019, all ages*. Available at: http://www.healthdata.org/Canada. (Accessed 13 November 2021).

Imamura, F., Micha, R., Khatibzadeh, S., Fahimi, S., Shi, P., Powles, J., & Mozaffarian, D. (2015). Dietary quality among men and women in 187 countries in 1990 and 2010: A systematic assessment. *Lancet Global Health, 3*(3), e132–e142.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666.

Just, D. R., Mancino, L., & Wansink, B. (2007). *Could behavioral economics help improve diet quality for nutrition assistance program participants?*.

Kirkpatrick, S. I., McIntyre, L., & Potestio, M. L. (2010). Child hunger and long-term adverse consequences for health. *Archives of Pediatrics & Adolescent Medicine, 164*(8), 754–762.

Kirkpatrick, S., & Tarasuk, V. (2003). The relationship between low income and household food expenditure patterns in Canada. *Public Health Nutrition, 6*(6), 589–597.

Lakdawalla, D., & Philipson, T. (2009). *The growth of obesity and technological change*. National Bureau of Economic Research.

Lakdawalla, D., Philipson, T., & Bhattacharya, J. (2005). Welfare-enhancing technological change and the growth of obesity. *The American Economic Review, 95*(2), 253–257.

MacLean, R. R., Cowan, A., & Vernarelli, J. A. (2018). More to gain: Dietary energy density is related to smoking status in US adults. *BMC Public Health, 18*(1), 365.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*(2), 159–179.

Newby, P. K., & Tucker, K. L. (2004). Empirically derived eating patterns using factor or cluster analysis: A review. *Nutrition Reviews, 62*(5), 177–203.

Nshimyumukiza, L., Lieffers, J. R. L., Ekwaru, J. P., Ohinmaa, A., & Veugelers, P. J. (2018). Temporal changes in diet quality and the associated economic burden in Canada. *PLoS One, 13*(11), Article e0206877.

Olstad, D. L., Campbell, N. R. C., & Raine, K. D. (2019). Diet quality in Canada: Policy solutions for equity. *Canadian Medical Association Journal, 191*(4), E100–E102.

Olstad, D. L., Nejatinamini, S., Victorino, C., Kirkpatrick, S. I., Minaker, L. M., & McLaren, L. (2021a). Socioeconomic inequities in diet quality among a nationally representative sample of adults living in Canada: An analysis of trends between 2004 and 2015. *The American Journal of Clinical Nutrition, 114*(5), 1814–1829.

Olstad, Dana Lee, et al. (2021b). Trends in Socioeconomic Inequities in Diet Quality between 2004 and 2015 among a Nationally Representative Sample of Children in Canada. *The Journal of Nutrition, 151*(12), 3781–3794.

Philipson, T. J., & Posner, R. A. (1999). *The long-run growth in obesity as a function of technological change*. National bureau of economic research.

Ricciuto, L. E., & Tarasuk, V. S. (2007). An examination of income-related disparities in the nutritional quality of food selections among Canadian households from 1986–2001. *Social Science & Medicine, 64*(1), 186–198.

Sadras, V. O., & Lawson, C. (2013). Nitrogen and water-use efficiency of Australian wheat varieties released between 1958 and 2007. *European Journal of Agronomy, 46*, 34–41.

Sanou, D., O'Reilly, E., Ngnie-Teta, I., Batal, M., Mondain, N., Andrew, C., Newbold, B. K., & Bourgeault, I. L. (2014). Acculturation and nutritional health of immigrants in Canada: A scoping review. *Journal of Immigrant and Minority Health, 16*(1), 24–34.

Schoufour, J. D., de Jonge, E. A. L., Kiefte-de Jong, J. C., van Lenthe, F. J., Hofman, A., Nunn, S. P. T., & Franco, O. H. (2018). Socio-economic indicators and diet quality in an older population. *Maturitas, 107*, 71–77.

Statistics Canada. (2017a). *Canadian community health survey - nutrition (CCHS)*.

Statistics Canada, H. (2017b). Reference Guide to understanding and using the data. Available at: https://www.canada.ca/en/health-canada/services/food-nutrition/food-nutrition-surveillance/health-nutrition-surveys/canadian-community-health-survey-cchs/reference-guide-understanding-using-data-2015.html. (Accessed 8 November 2021).

Statistics Canada, S. C. (2017c). Census profile, 2016 census - Canada. Available at: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=PR&Code1=01&Geo2=&Code2=&SearchText=Canada&SearchType=Begins&SearchPR=01&B1=All&TABID=1&type=0. (Accessed 13 November 2021).

Templ, M., Gussenbauer, J., & Filzmoser, P. (2020). Evaluation of robust outlier detection methods for zero-inflated complex data. *Journal of Applied Statistics, 47*(7), 1144–1167.

Unnevehr, L., Eales, J., Jensen, H., Lusk, J., Mccluskey, J., & Kinsey, J. (2010). Food and consumer economics. *American Journal of Agricultural Economics, 92*(2), 506–521.

Ver Ploeg, M., & Rahkovsky, I. (2016). *Recent evidence on the effects of food store access on food choice and diet quality*.

Yeung, C. W., & Thomas, S. (2012). *Income Imputation for the Canadian Community Health Survey*. Ottawa, ON: Statistics Canada.